Constraint Based Event Recognition for Information Extraction

Jeremy David Macdonald Crowe

Ph.D.
University of Edinburgh
1997

Abstract

A common feature of news reports is the reference to events other than the one which is central to the discourse. Previous research has suggested Gricean explanations for this; more generally, the phenomenon has been referred to simply as "journalistic style". Whatever the underlying reasons, recent investigations into information extraction have emphasised the need for a better understanding of the mechanisms that can be used to recognise and distinguish between multiple events in discourse.

Existing information extraction systems approach the problem of event recognition in a number of ways. However, although frameworks and techniques for black box evaluations of information extraction systems have been developed in recent years, almost no attention has been given to the evaluation of techniques for event recognition, despite general acknowledgment of the inadequacies of current implementations. Not only is it unclear which mechanisms are useful, but there is also little consensus as to how such mechanisms could be compared.

This thesis presents a formalism for representing event structure, and introduces an evaluation metric through which a range of event recognition mechanisms are quantitatively compared. These mechanisms are implemented as modules within the CONTESS event recognition system, and explore the use of linguistic phenomena such as temporal phrases, locative phrases and cue phrases, as well as various discourse structuring heuristics.

Our results show that, whilst temporal and cue phrases are consistently useful in event recognition, locative phrases are better ignored. A number of further linguistic phenomena and heuristics are examined, providing an insight into their value for event recognition purposes.

Acknowledgements

As the next page proudly states, I composed this thesis entirely myself. What it omits to say is that I would probably *still* be composing this thesis were it not for the subtly insistent tones of my supervisor Chris Mellish. More importantly, without the benefit of his experience, advice and unfailing encouragement, I shudder to think what form the composition would have take on. I thank him sincerely.

Many thanks are also due to my second supervisor Marc Moens for reading earlier drafts of this thesis, and for making suggestions that reveal a sometimes frighteningly conscientious attention to detail.

I am grateful to the Economic and Social Research Council, who provided me with funding, and to the Artificial Intelligence Department at Edinburgh University, who have provided me with a roof over (and, on one exciting occasion, even *in*) my head over the last four years, and who have made it possible for me to attend conferences overseas.

A large number of other people have been of help during all stages of this thesis, including anonymous conference reviewers and colleagues both within the AI Department and further afield. Although it's not strictly true that the denizens of E17 have accelerated the completion of my work, they have at least proved that procrastination can be a lot of fun.

Thanks also to my family, Flora, David, Jonny and Rob for, in no particular order, emotional support, financial support, and suggesting that I take a break during the final weeks of writing up and get my paragliding licence — you know who you are.

Finally, eternal thanks to Vanessa for being both the ultimate distraction and the best reason to finish.

Declaration

I hereby declare that I composed this thesis entirely myself and that it describes my own research.

Jeremy David Macdonald Crowe
Edinburgh
April 29, 1997

Contents

List of Figures

List of Tables

Frequently used abbreviations

IE	information extraction
ER	event recognition
TA(M)	temporal analysis (module)
LA(M)	location analysis (module)
CPA(M)	cue phrase analysis (module)
EM	event manager
MUC(-n)	(the n^{th}) Message Understanding Conference

And now for something completely different.

Newsreader, [Chapman *et al.* 70]

Chapter 1

Introduction

> The modern age has a false sense of security because of the great mass of data at its disposal. But the valid issue is the extent to which man knows how to form and master the material at his command.
>
> Johann Wolfgang von Goethe (1749 – 1832)

The ability to understand natural language is one of the ultimate goals of Artificial Intelligence. The rest of the world, however, cannot afford to wait. In the meantime there are clearly defined information management problems that AI is already starting to solve. New fields of research, such as Language Engineering in general, and Information Extraction in particular, hint at a rising sense of optimism about what can be achieved in the short term.

Recognised obstacles lie in the way of some of the most promising of these technologies, however. In part due to the sheer scale of the task, potential solutions to a widely acknowledged problem in information extraction have become hard to identify. This thesis penetrates the opacity that black box evaluations of Information Extraction have suffered from, and casts light on some promising ways forward.

1.1 Information extraction

The above quotation from Goethe has never been more relevant than it is today. As ever increasing amounts of data are stored and transmitted in electronic form, information processing needs grow accordingly. Information extraction (IE) can be seen as the task of finding useful information in a collection of texts and processing that information into a format that can be inserted into a database. Although information extraction is a new field of research, its roots can be traced back to two main sources.

In the late 1970s and early 80s, a group of researchers at Yale University, led by Roger Schank, were producing interesting results in the field of story understanding [Schank & Abelson 77] [Cullingford 78] [Wilensky 78] [DeJong 79] [Lebowitz 80]. Their systems attempted to produce *full coherent interpretations* of quite difficult texts. Although they ran into obstacles in scaling up their systems beyond anything but quite restricted domains, their research resulted in some novel theories of memory organisation [Schank 80] and an unspoken challenge to the NLP community to do better.

At roughly the same time, Naomi Sager at New York University was taking a more goal-oriented approach to text processing [Sager 78] [Sager 81], with the aim being simply to format documents in such a way that *predetermined elements* could be encoded into a database. Sager certainly didn't claim to be doing natural language *understanding*; instead, she used the more modest (and certainly less emotionally loaded) term *information formatting.* Rather than relying on large, fragile knowledge sources to attempt a complete analysis of the text, Sager was using application-specific rules to convert idiosyncratic documents into regularised, explicit representations. Furthermore, whereas the Yale group's work was technology-driven, in so far as the research focus was on the methods used, Sager's approach was much more problem-driven, the focus being on providing a solution to an existing clearly defined problem (that of processing radiology reports).

Following a relative hiatus during the mid 80s, interest in the field of message understanding revived, largely due to the awakening of the US Government to the mounting problem of analysing online information, the growth of which was by now starting to pick up speed. Under the organisation of the (D)ARPA sponsored Tipster initiative, which was already funding evaluation-oriented research into information retrieval, the Message Understanding Conferences (MUC) [DAR91] [DAR92] [ARP94] [ARP96] have concentrated research in robust text processing through the development of rigorous evaluation metrics and a forum in which to compare solutions and scores (though not necessarily in that order).[1]

Although the title of the evaluation may suggest otherwise, MUC is not really about message understanding. Few, if any, of the systems entered in the recent evaluations aim to produce a complete interpretation of the documents. This is hardly surprising, given the heterogeneous nature of the input text and, just as importantly, the regularised target representations for which systems are scored. The MUC evaluations are about information extraction, and although the motivation and inspiration may have originated in the sort of systems being developed at Yale in the late 70s, current approaches have more in common with Sager's information formatting.

1.2 Event recognition

Information extraction technology has been applied to many genres of text (about which we shall say more later). One of the most popular genres is news texts, for reasons that are both practically motivated (news text is widely available in electronic form, and there are a wide range of potential applications, about which also more later) and theoretically motivated (news texts contain challenging linguistic properties, whilst remaining stylistically fairly neutral). The corpus we have adopted falls squarely into this genre.

One of the properties of news texts is their tendency to describe more than one event, often for purposes of comparison or background information (see section 2.3). Defining precisely what is meant by an *event* is difficult. Our understanding of an event is perhaps best demonstrated through the use of an example. The following text, though constructed for the purposes of illustration, is representative of the documents we are

[1] Examples of MUC texts and output templates are contained in appendix E.

interested in.

```
TST1-DEMO-0001

BOGOTA, 4 FEB 90 (ACAN-EFE) -- [TEXT] A BOMB EXPLODED YESTERDAY IN
DOWNTOWN ARACATACA. POLICE IN BOGOTA SAY THAT THE JPF WERE RESPONSIBLE
FOR THE ATTACK. SEVERAL BUILDINGS WERE DAMAGED IN THE BLAST.

IN A SIMILAR INCIDENT TEN DAYS AGO, THE JPF ATTACKED AN ARMY INSTALLATION
IN THE TOWN OF RIVERA. THE GUARDS SHOT TWO TERRORISTS.

JPF GUERILLAS ARE KNOWN TO FREQUENT THE TOWN.

SATURDAY'S ATTACK OCCURRED AT AROUND 1725 LOCAL TIME, AND MAY HAVE BEEN
TIMED TO DISRUPT RUSH HOUR TRAFFIC.
```

We would claim that this text describes two events, one of which is a bomb attack that took place "yesterday" in Aracataca, and the topic of the first and last paragraphs, and the other, a related incident in another town that resulted in two terrorists being shot, the focus of the second paragraph.

Clearly it could be argued that the above text actually describes far more "events" than this, such as an explosion, a report by the police in Bogota, the infliction of structural damage and so on. The issue here is one of granularity of information. In a sense what we are interested in is the *smallest number of coherent events* covering *as much of the text* as is possible, in this case the two picked out above. This is the level that information extraction systems are required to produce. However, as we shall see in chapter 2, in the process of attempting to reach this goal, systems first have to identify "events" at the finest level of granularity. As it happens, the events we are referring to in the above text exist at the paragraph level. Clearly this does not necessarily have to be the case — they could be at anything from the phrasal level up to the level of the whole document.

For information extraction systems, recognising that a text contains multiple events (such as the example text above) is crucial. Even if deep linguistic approaches such as full text parsing and semantic interpretation were able to scale up to the IE task, problems of coreference in the entity-rich genre of news texts would remain. For systems employing more shallow approaches, the motivation to distinguish between multiple events is even greater, as they are also more exposed to problems in determining relationships between entities. Without an awareness of the number and distribution of events in a text, systems run the very real risk of incorrectly relating unassociated entities, resulting in the potential confusion and/or omission of information.

The need for a better understanding of the ways in which new events are introduced in a discourse is widely accepted among those working in the field of information extraction. One of the most consistently successful teams of system developers has cited the distinction of multiple events as one of the three "*major areas for future improvements in MUC-like tasks*", [Krupka *et al.* 92]. Furthermore, it seems a widely held belief within the information extraction community that "*identifying portions of text that describe different domain events* [is] *crucial not only for* [information extraction] *but for text understanding in general*" [Iwańska *et al.* 91]. Those researchers most intimately involved in evaluation of IE systems have pointed out the "*inadequacy of current approaches to determine when and how to combine information from multiple sentences into a single, coherent representation*" [Sundheim 92]. More than a year later, the conclusion of a workshop on discourse issues of IE concluded that "*discourse processing remains one of the major outstanding issues in data extraction*" [Ayuso 93]. Backing this up, an exploratory study reported by Hirschman [Hirschman 92] states that there is an observed "*degradation ... as the information distribution* [becomes] *more complex*", as Hirschman's own findings seem to confirm. For IE, therefore, event recognition is far more than simply an area of theoretical interest; rather, it is a practical problem of significant importance.

1.2.1 Event recognition and event distinction

It may seem that we have been using the terms event *recognition* and event *distinction* in an interchangeable manner. However, although the two terms are inextricably related, the difference between the two is important. The term *event distinction* can be used to describe the process of identifying an *event shift* between two (not necessarily adjacent) parts of a text. However, it does not tell us what those units *are* that correspond to different events, just where there is a difference. In cases where the (theoretically unitless) areas of text are adjacent, event distinction is analogous to *boundary* detection [Morris & Hirst 91] [Grosz & Sidner 86].

On the other hand, *event recognition* ties together textual elements that refer to the same event — in other words, it is a process for defining event *units*. It does not in itself identify boundaries. Of course, if we assume that every element (sentence, clause etc.) belongs to *an* event, then distinction and recognition can be seen as two different approaches to the same problem, i.e. attributing textual elements to events. In practice, however, IE systems tend to maintain the dichotomy, partly because (depending on the application) not all of the text may be relevant, but more importantly because no single technique lends itself to performing either recognition or distinction throughout the text. The properties of a text that indicate "sameness", i.e. the continuation of an event, seem to be different to those that signal "difference", i.e. the transition between events, as we shall see in the next chapter.

Having said this, we will often want to refer to both event recognition and event distinction simultaneously in this thesis. Because of this, we shall use the term *event recognition* in a general sense to mean both the recognition and distinction of events. Where we wish to imply the specific meaning of the term, we shall make this clear.

1.2.2 Event recognition and domain type

As we stated previously, the genre with which this thesis is concerned is that of news texts. In particular, we are using the Latin American terrorist attack corpus that constitutes the MUC-3/4 development and test corpus. Our reasons for using this corpus are simple: terrorist attacks, by their very nature, are sudden and quite instantaneous in nature, and as such are described in terms of events (as opposed to states) and, importantly, are often described in the context of previous (and related) events. This phenomenon of describing multiple events (whether related or not) within the bounds of a single text is clearly not restricted to terrorist attacks. The Air Activity Corpus, which contains military messages describing the movement of aircraft, appears to be very similar in this respect [Stalls *et al.* 90]. Rather, it is a technique widely used by speakers in general — reasons for this are suggested in section 2.3.

Clearly there are discourse types where this concept of "event" is less tangible. Bank telexes, for example, may contain multiple banking events (e.g. credits, withdrawals), but not for the same reason that news texts contain multiple events. The events in the domain of financial transactions are also more predictable, and form part of a closed set of possible actions. (Nevertheless, challenging IE tasks have been undertaken in the field of finance, for example [Young & Hayes 85].)

1.3 Aims of the thesis

This thesis has several aims, the most important being to provide some indication of the sort of linguistic phenomena that provide the best clues as to the event structure of a text. Deciding what constitutes the "best" requires us to be able to evaluate event recognition — something that has only been rarely done previously, and in those cases quite informally.[2] However, in order to carry out an evaluation, we must have something to evaluate, which requires a suitable representation for describing event structure. The theories embedded in these aims therefore need to be implemented within a configurable framework.

As we suggested earlier, shallow approaches to IE (and other areas of NLP) have yielded encouraging results. Our aim is therefore to test the feasibility of using shallow techniques for event recognition, through the establishment of a cognitively and mathematically sound discourse representation formalism and evaluation methodology.

1.4 Thesis outline

The thesis can be viewed as having three parts. The first part, which this chapter begins, continues with chapter 2, which surveys the field of information extraction in general, and the task of event recognition in particular. Also presented is a broad classification of current event recognition techniques, with illustrations drawn from a number of recent IE systems.

The second part of the thesis includes chapters 3 through to 7, which discuss in detail

[2] Although the evaluation of IE systems in general is at a formal level, it has tended to take the form of black box evaluations. Furthermore, given the typically monolithic nature of IE systems, in particular the older ones, it is often extremely difficult, if not impossible, to judge the performance of individual components in these systems. This is a problem that the IE community is currently striving to resolve, for example through the GATE (Generalised Architecture for Text Engineering) framework [Cunningham *et al.* 95] proposed by IE researchers at Sheffield University.

the main contribution of the thesis. Chapter 3 introduces our approach to ER in terms of formalisms and mechanisms. In the case of the former, we describe various properties of the corpus we have used (the input) and the clause-event grids with which we represent the discourse structure (the output). The mechanisms introduced include, at a general level, the system we have built to test our hypotheses about text segmentation, CONTESS. In more specific terms, we also introduce its component parts, the three analysis modules and the Event Manager.

We then present the linguistic phenomena that CONTESS is concerned with, and the components that have been developed to process them. Chapter 4 describes our approach to temporal analysis, beginning with a discussion of the nature of temporal information in discourse and continuing by showing how temporal phrases may be defined and identified. The Temporal Analysis Module is then described in detail, and fully illustrated using a short example text.

Chapter 5 describes the second of the three analysis modules, the Locative Analysis Module. The chapter begins by discussing the ways in which locative information is presented in discourse, and in particular in news texts. A working definition of locative phrases is described, followed by a detailed presentation of the Locative Analysis Module, together with an illustration of its use.

Chapter 6 introduces the last, and in this case the least, of the three analysis modules. Cue phrases have previously been used as an aid to event distinction, and in this chapter we look at some of the earlier approaches that have been adopted. The Cue Phrase Analysis Module is a simple goal-directed system for processing a subset of cue phrases, and is also discussed here.

The final chapter in this part of the thesis describes the Event Manager, the component that receives input from the above three analysis modules and produces an output clause-event grid (or multiple grids). This chapter focuses on the dual identification roles of the EM — that of identifying both the discourse units (i.e. event recognition), and the relationship between those units (i.e. event distinction). The suite of heuristics that the EM uses to propose the output grids are also described and, again, the Manager is illustrated using the example text first encountered in chapter 3.

The third part of the thesis concerns evaluation of event recognition. Chapter 8 describes the various aspects of evaluation that we are interested in, i.e. the evaluation of our discourse structure representations, the clause-event grids; the evaluation of CONTESS as a text segmentation tool; a more detailed evaluation of the components within the system, in order to gain insight into the relative benefits of processing particular linguistic phenomena; and an evaluation of the level of agreement that there exists between human coders with respect to both our representation formalism and the event recognition task.

Chapter 9 describes the techniques that can be used to summarise the results of these evaluations, and discusses the findings, concluding that while the Temporal Analysis Module (and by extension temporal information) and, to a lesser extent, the Cue Phrase Analysis Module, constitute useful tools for event distinction, the Locative Analysis Module appears to be quite unsuited to the task. Further inferences are drawn from the observed performance of the heuristics used by the Event Manager. The results and implications of the agreement test carried out with human coders are also discussed.

Finally, chapter 10 concludes the thesis with a summary of the relative performance of the individual components of CONTESS, and of the system as a whole. Also discussed are the strengths and weaknesses of the approach we have taken, together with some pointers to areas that we believe warrant further research.

Extensive appendices present an example of CONTESS processing a news text; detailed tables of results for each system evaluation conducted; the correlation subcorpus used to measure agreement between coders, together with a graphical visualisation of the agreement scores for each document; and a short list of sample documents referred to during the text.

Chapter 2

Literature survey

2.1 Introduction

This chapter provides a general background on information extraction in general, and event recognition in particular, and discusses some of the issues that have been raised by researchers confronting the problem of processing multi-event texts. We present a broad classification of event recognition techniques, and examine in some detail approaches representative of these types.

2.2 Information extraction

Information extraction can be seen as a natural progression from work in text understanding. In the late 1970's and early 1980's, researchers at Yale[1] led by Schank made significant steps forward in the robust processing of "real" pieces of text, with various systems pioneering the use of highly event-specific scripts [Schank & Abelson 77].

2.2.1 Scripts

Scripts use clustered links of conceptual dependency [Schank 72] relations to represent stereotypical events as a sequence of time-ordered sub-events (scenes), together with various conditions, results and the required and optional roles and props that feature in an event. For example, a trip on a subway train involves roles such as groups of

[1] See [Lehnert 94] for a lively and revealing insight into the early years of text understanding.

passengers, a conductor and a subway organisation; props such as tickets, money, trains and seats; and scenes such as entering a subway station, buying a ticket and travelling on the train.

In rendering explicit the causal relationships between subevents, scripts allow a text understanding system to make inferences about connections between entities and actions. Consequently, such systems are able to interpret texts where the links between actions are left unstated. As this is very often the case in real texts, this is a powerful position to be in.

As well as allowing a text understanding system to build a single coherent interpretation of a discourse, scripts provide an elegant way of focusing on unusual events (i.e. events that fail to correspond to the predicted chain of causal relations). Cullingfords's SAM [Cullingford 78] employs this kind of reasoning.

2.2.2 Scripts and events

Lebowitz's IPP system [Lebowitz 80], also developed at Yale, uses scripts as a means of deriving generalisations about stereotypical events based on interpretations of several news articles. Entities or relationships that are absent from the incident-specific script and yet appear in more than one news article are proposed as generalisations. Although IPP doesn't have to interpret texts containing references to multiple distinct events, it does at least process multiple events, and so has to deal with the problem of merging entities from different events.

When presented with a series of newswire stories on Middle East terrorist attacks, for example, IPP formed the generalisation that *terrorists responsible for bombings in Israel usually escape*. On the other hand, given an insufficiently large number of texts it will infer generalisations such as *terrorist actions in India always result in two deaths.* Lebowitz notes, however, that people often do make these sorts of inferences when provided with information that fits such a generalisation.

In so far as IPP needs to know what can correspond to, say, a kidnapping event, it does have a technique for determining what elements of a text correspond to a continuation of an event – in other words, it performs a simple form of event recognition. However, the event recognition is not in a multiple event context; it is only used as a means of structuring sub-events within more general events.

For example, and using Lebowitz's terminology, the most general type of events (S-MOPs, for Simple Memory Organisation Packets) are divided into categories like *extortion*, which consist of more specific event types (spec-MOPs) such as *kidnapping* and *hijacking*. Although these events contain many of the same properties (hence their inclusion under the same S-MOP), they will also have some events that are particular to each. Finally, action units describe "concrete events, such as shootings, people being wounded, and hostages being released." Figure 2.1 (adapted from [Lebowitz 80]) illustrates this taxonomy.

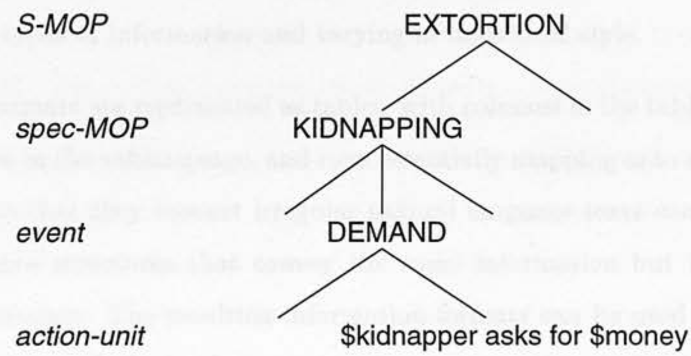


Figure 2.1: Event structure in IPP

2.2.3 From text understanding to information extraction

Scripts were never envisaged as a practical method of guiding text understanding. The hand-coded knowledge-intensive scripts were not only difficult and tedious to write, but resulted in systems that were vast, slow and fragile [Cullingford 86] [Schank 86]. DeJong's news skimmer FRUMP [DeJong 79] partly solved the time problem by using "sketchy scripts", working on the basis of initially identifying the topic of the news article and then, if it had a sketchy script for that topic, searching for the domain-specific elements of that article. The shallow, directed stance that FRUMP took is suggestive of the approach taken by many of today's information extraction systems.

Despite the top-down guidance to text understanding that scripts afforded the Yale systems and others like them, the approach was still very much one of viewing the text as a *source* from which material (in the form of the roles, props and causal relations in a text) was to be gleaned.

Another approach views the information to be extracted from a text as the *goal* of text processing. The difference is a subtle one, but more than simply one of emphasis. It means that whereas the former approach attempts to extract as much information from the text as is possible, the latter only extracts that which is required to fulfill the information goal.

Naomi Sager's work on *information formatting* [Sager 78] is an example of this. Although she was working within far more restricted domains[2], the texts that were being processed (clinical reports) were nonetheless quite complicated, containing a considerable range of types of information and varying in individual style.

Information formats are represented as tables, with columns in the table corresponding to word-classes in the sublanguage, and rows essentially mapping onto sentences. Their main benefit is that they convert irregular natural language texts containing implicit information into structures that convey the same information but in a regularised and explicit manner. The resulting information formats can be used for information extraction and compilation tasks.

The goal in Sager's work was therefore certainly not deep text understanding. Rather, it was the extraction of specific well-defined items of information (those items for which a sublanguage description and formatting rules could be built).

A more recent and vivid shift from text understanding to information extraction can be seen in the systems fielded by Jerry Hobbs and others at SRI in recent evaluations at the ARPA-sponsored Message Understanding Conferences [Sundheim 91] [Sundheim 92] [Chinchor *et al.* 93].

The system entered for the third evaluation, TACITUS [Hobbs *et al.* 91], is very much a text understanding system. It attempts to analyse every word of the sentences in its (highly complex) input texts[3], and interprets the resulting logical forms within an abductive inference framework. TACITUS achieves high scores [Hobbs 91], but takes 36

[2] Sager doubts information formatting would be applicable outside of restricted domains [Sager 81].

[3] A relevance filter is used to detect sentences that are relevant for the extraction task.

hours to process 100 messages.

The following year the SRI team entered FASTUS [Hobbs *et al.* 92a] which, despite its acronym[4], is an information extraction system rather than a text understanding system. It achieved equally impressive results in the evaluation, yet without any of the deep natural language capabilities of its predecessor. In recognising the presence of a "pre-defined ... rigid target representation" and the irrelevance of the writer's goals in writing the text [Hobbs *et al.* 92b], FASTUS was able to use a finite state approximation of a highly restricted grammar and rudimentary semantic processing and still arrive at very respectable levels of performance in terms of both effectiveness and speed.

Having looked at the difference between text understanding and information extraction, and seen some examples of both types of text processing, we now turn to the issue of recognising and distinguishing between events in information extraction.

Although early news understanding systems were processing real (albeit very short) documents, texts usually referred to just one event. Given the nascent state of text understanding at that time, this was certainly a fair limitation to impose on the material. For practical information extraction purposes, however, the ability to handle multiple distinct events in a piece of text is of huge importance. This is backed up by many researchers working on IE systems, as we saw in the introduction to this thesis. In the next section, we will look at previous work on the analysis of multi-event discourse and, in particular, news texts.

2.3 Discourse structure of news reports

Iwańska has described two types of discourse structure in news text, which she terms *shared information* (SI) structures and *embedded* (E) structures. These are illustrated in figure 2.2 (adapted from [Iwańska 93]). SI-structures are defined as containing a *shared information before* (SIB) segment containing information that is shared by a set of entities described later in the text; a number of segments following the SIB segment that individually describe these entities; and a *shared information after* (SIA) segment following these segments that contains information relating to the previously

[4] FASTUS is a permuted acronym of Finite State Automaton Text Understanding System.

introduced entities. The SIB segment sometimes corresponds to the introduction of the text [vanDijk & Kintsch 83], and the SIA segment to the conclusion. Iwańska claims that texts containing both SIB and SIA segments are rare, while texts with either one or neither of the shared segments are very common.

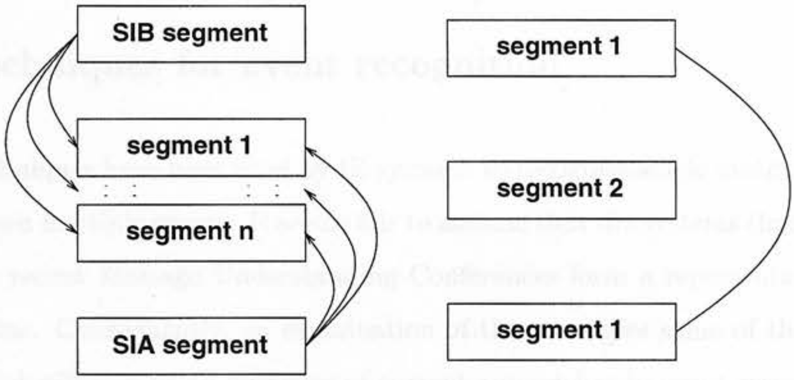


Figure 2.2: Iwańska's SI structures (left) and E structures (right)

E structures, which Iwańska reports are much more common than SI structures, contain embedded segments describing other (often somehow related) entities. She also points out that, while previous research (for example, that of Grosz and Sidner [Grosz & Sidner 86]) has focused on shorter, recursive discourse structure, news texts typically have a much flatter structure. The primary reason for this, it would seem, is due to the function of news texts — i.e., to convey factual information with minimal misunderstanding. This may also be why combinations of SI and E structures are not found in news texts.

Moens [Moens 92] suggests a Gricean explanation for structural aspects of factual texts, claiming that the maxim of *Relevance* dominates at the document level (i.e. above the paragraph level), whereas below this the maxim of *Manner* imposes an orderly sequence unless explicitly marked otherwise.

Dirk [Noel 86] adopts a Rhetorical Structure Theory [Mann & Thompson 88] analysis, and finds that whereas instructional texts usually use *motivation*, *enablement* and *solutionhood* as relations between clauses, news texts are more likely to use *elaboration*, *background* and *addition*. Headlines and summaries (usually the first and last paragraphs in the text) are nearly always classed as *restatement*. Although this is hardly a surprising finding, it nonetheless serves to reinforce the notion that news texts possess idiosyncratic properties that might be exploited in a practical event recognition mechanism. In the next section, we will look at some of the ways in which current information extraction systems perform event recognition.

2.4 Techniques for event recognition

Various techniques have been used by IE systems to recognise single events and distinguish between multiple events. It seems fair to assume that the systems that have taken part in the recent Message Understanding Conferences form a representative sample of IE systems. Consequently, an examination of the strategies some of these systems have adopted will give us an overview of current approaches to event recognition.

One point that emerges is that not all systems explicitly perform event recognition and distinction. Indeed, most systems opt solely for explicit event recognition. Some of the most successful systems, however, also make use of explicit event distinction.

Table 2.1 provides a summary of techniques, with columns giving (from left to right) the name of the system; the group responsible; the version of system in terms of the evaluation it was fielded at; the recognition technology used; the status of explicit event recognition in the system; and the status of explicit event distinction.

2.4.1 Frame merging

The most popular method of recognising single events and distinguishing between multiple events is usually known as frame (or template) merging. Systems using this approach typically create frame-like data structures for each clause (or sentence) based on some form of semantic representation. These frames are then checked for compatibility (either incrementally or as a single process after analysis), where compatibility is typically determined by event type (e.g. bombing, arson), location and time. Other properties, such as compatible targets, are sometimes specified.

Successful frame merging is highly dependent on accurate and (at least locally) full semantic analysis which, in an application such as IE, is very hard to achieve. In checking

System	Group	Version	Technique	ER	ED
Alembic	Mitre Corporation	MUC-4	FM-	√	×
Codex	Advanced Decision Systems	MUC-3	FM	√	×
Plum	BBN Systems and Technologies	MUC-3	FM	√	×
Plum	BBN Systems and Technologies	MUC-4	FM	√	×
TIA	GTE Government Systems	MUC-3	FM	√	×
DBG	Language Systems Inc	MUC-3	FM+	√	×
DBG	Language Systems Inc	MUC-4	FM	√	×
Inlet	McDonnel Douglas	MUC-3	FM?	?	?
Paktus	PRC	MUC-3	FM	√	×
Paktus	PRC	MUC-4	FM	√	×
Solomon	SRA	MUC-4	FM	√	×
Fastus	SRI International	MUC-4	FM	√	×
N/A	Synchronetics	MUC-3	FM-	√	×
Ictoan	University of Maryland	MUC-4	FM?	?	?
Link	University of Michigan	MUC-4	FM	√	×
N/A	UNL/USL	MUC-3	FM	√	×
Snap	Univ Southern California	MUC-4	FM	√	×
Texus	McDonnel Douglas	MUC-4	FM+	√	×
Proteus	New York University	MUC-3	FM+	√	×
Proteus	New York University	MUC-4	FM+	√	×
N/A	Paramax Systems	MUC-4	FM+	√	×
N/A	Unisys	MUC-3	FM+	√	×
Circus	University of Massachusetts	MUC-4	FM+	√	×
MucBruce	NMSU/Brandeis	MUC-4	PS FM	√	√
NLToolset	General Electric R&D	MUC-3	PS FM PoS	√	√
NLToolset	General Electric R&D	MUC-4	PS FM PoS	√	√
Shogun	General Electric R&D/CMU	MUC-4	PS FM PoS	√	√
TTS	Hughes Research Labs	MUC-3	TG FM	√	√
TTS	Hughes Research Labs	MUC-4	TG FM	√	√-
Interpretext	Intelligent Text Processing	MUC-3	PoS	×	√
Circus	University of Massachusetts	MUC-3	FM PoS	√	√
Tacitus	SRI International	MUC-3	AM	×	×

Key: FM frame merging; PS presegmentation; PoS postsegmentation; TG topic grouping; AM abductive merging

Diacritics: ? unclear; + enhanced; - reduced

Table 2.1: Summary of event recognition and distinction techniques in MUC-3 and MUC-4

for compatibility between two slot/filler pairs, accurate interpretation of entities (for example, organisations in the case of NP resolution) can make the difference between correctly recognising two references to the same entity (and therefore compatibility) and incorrectly assuming references to two distinct entities (i.e. incompatibility). However, the intuitive attraction of the technique is clear, as it represents implicitly our assertion in section 1.2 that the set of events worth recognising corresponds to the minimum number of possible events.

An example of frame merging is shown in figure 2.3 (adapted from [Hobbs *et al.* 92b]). Slot names correspond to incident type, perpetrator, confidence in incident, and the human target.

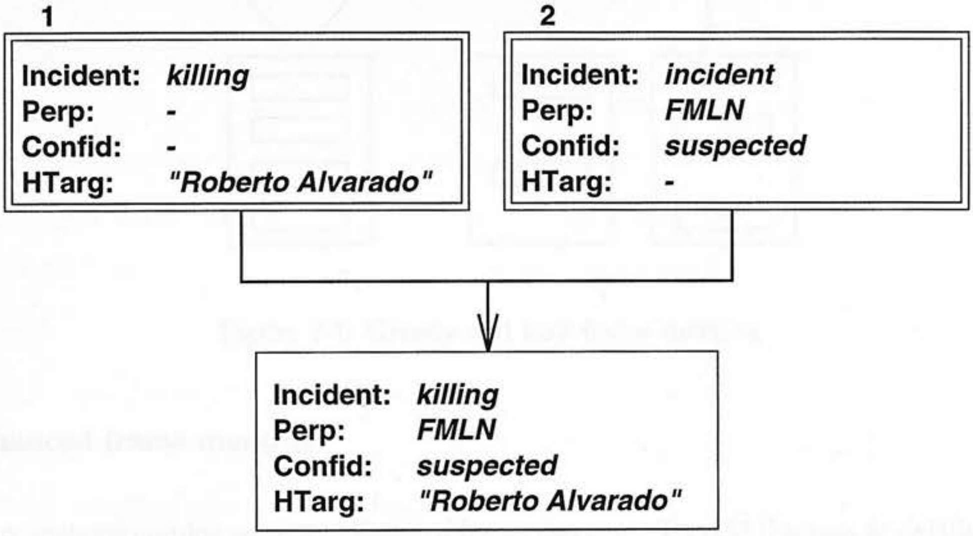


Figure 2.3: Example of frame merging

In this example, the first frame is produced by the clause "*killing of Attorney General Roberto Alvarado* ", while the second is generated by "*Salvadoran President-elect Alfredo Cristiani ... accused the Farabundo Marti National Liberation Front (FMLN)*". As "killing" is more specific than "incident", the former becomes the filler for the incident slot. The rest of the frame is filled by taking the union of the other slots. BBN's PLUM [Weischedel *et al.* 91], GTE's TIA [Dietz 91], PRC's PAKTUS [Loatman 92] and SRI's FASTUS [Hobbs *et al.* 92b] are representative of this approach.

Greedy and lazy frame merging

The incorrect merging and non-merging of frames has been described respectively as *greedy* and *lazy* merging [Hirschman 92], and is illustrated in figure 2.4. In the case that merge A of clauses C1 and C2 is correct, merge B corresponds to a lazy merge. If, on the other hand, merge B were correct, merge A denotes a greedy merge. Greedy merging results in missed events and, possibly, spurious slot fills, whilst lazy merging leads to spurious events and missing slot fills.

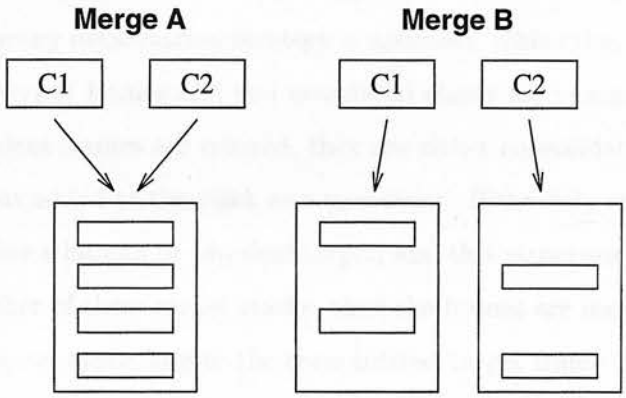


Figure 2.4: Greedy and lazy frame merging

Enhanced frame merging

Many systems employ enhanced forms of frame merging. TEXUS [Meyers & deHilster 92] uses pragmatic rules to merge events based on location so that, for example, clauses describing attacks on buildings are linked with those clauses describing the presence of people inside the buildings. The IE system developed by Unisys [Weir *et al.* 91] and, subsequently, Paramax Systems [Weir & Silk 92], merges events of the same incident type if there is a significant overlap in the text regions they occupy, or if they share human or physical targets. PROTEUS [Grishman *et al.* 92] behaves similarly, attempting to merge frames generated from within the same sentence. It also attempts to merge *effect* frames when they follow *attack* frames, as in the following example.

The JPF attacked the village. Two civilians were injured.

Lehnert *et al*'s early version of CIRCUS [Lehnert *et al.* 91] used a rule-based frame merging approach that soon became unmanageable [Lehnert *et al.* 95] as it grew to contain nearly 170 rules. Motivated by this problem, the group developed a new technique for their system in the 1992 MUC evaluation, termed *memory-based consolidation* (MBC). Memory-based consolidation is functionally isomorphic to frame merging in that it involves the merging (under certain domain-specific compatibility constraints) of data structures that represent individual (or, after merging, multiple) sentences in a text.

The MUC-4 CIRCUS's [Lehnert *et al.* 92] merging process is memory-based in the sense that a specific memory organisation strategy is assumed. This takes the form of a stack of incident (event-type) frames and two associated stacks for human and physical targets. As new incident frames are created, they are either consolidated into an existing one on the stack, or added to the stack as a new event. If the data structure created by a sentence describes a human or physical target, and this structure is compatible with one already on either of these target stacks, then the frames are merged. Furthermore, the incident frame corresponding to the consolidated target frame is moved to the top of the stack, so that it will be the first incident to be considered in the next round of merging.

Some IE systems go beyond the level of these relatively simple enhancements, and use additional pre- or post-analysis[5] segmentations – usually in combination with frame merging, but sometimes in isolation.

2.4.2 Presegmentation

Successful IE systems perform at least some form of linguistic analysis. Certain groups have made use of mechanisms for suggesting discourse-level segmentations of texts *before* such linguistic analysis has taken place. Early awareness of the event structure of a text offers computational savings later on – primarily in reference resolution of NPs and pronominal anaphora.

[5] *Analysis* rather than *parsing* – not all IE systems operate on syntactic structures.

NLTOOLSET

The NLTOOLSET system designed at General Electric [GECS 92] [Krupka *et al.* 91] [Krupka *et al.* 92] uses presegmentation to distinguish between events as part of its preprocessing. The Discourse Processing Module [Iwańska 91] is a heuristically-driven program that, working in conjunction with the postsegmentation component TRUMPET (see section 2.4.3), makes use of key phrases, definiteness and text quantity measures in postulating segments of text that denote distinct events. An important aspect of the key phrases component is its decomposition of phrases into *primary* and *secondary activators*. Primary activators[6] are incident type-related[7] phrases that are used to determine both the start of a new event (in the case where the previous event belongs to a different incident type) and the continuation of an event. This simple approach to event recognition has one obvious disadvantage — it is not necessarily the case that, given two distinct events in a text, each will belong to a different incident type. Secondary activators are used to decide the closing boundaries of events, e.g. 'serious condition'. It is not clear exactly how these phrases are arrived at, nor is it evident how large the set of secondary activators is. Essentially, though, the activators are used to determine the incident type of fragments of text, with primary activators roughly performing what we have termed event *distinction*, and secondary activators performing *recognition*.

The Discourse Processing Module (DPM) uses a multi-pass algorithm. The individual stages are actually very complex indeed, containing many conditions and caveats. However, it can be very roughly summarised as follows:

1. First pass segmentation of story
   - compute primary and secondary activators for document
2. Further splitting of segments
   - examine correlation between cue phrases (e.g. 'meanwhile'), activators and definiteness of activator patterns

[6] Activators are described using a dedicated regular language [Jacobs *et al.* 91].

[7] In the case of MUC-3/4, the incident types are *arson*, *bombing*, *forced work stoppage*, *hijacking*, *kidnapping*, *robbery* and the more general *attack*.

- look for explicit signallings of new events

3. Merge segments

    - check for incident type subsumption relations
    - check for reappearances of proper names
    - merge non-sequential segments under specific activator situations

As is typical of work in this area, the success of DPM's approach is very hard to judge. Iwańska claims that it is fast and that it "*performs well on the large MUC-3 corpus*". Some examples of the algorithm working with various degrees of success are given, and they allow us some insight into the strengths and weaknesses of this approach.

For a system that has no explicit knowledge about discourse structure, it performs quite well on a range of documents. One of its strengths is that it only cares about the definite/indefinite nature of incident-loaded phrases, so that 'a bomb' is a useful indicator of a new event, whereas 'a building' isn't. On the other hand, it is unclear whether locative NPs such as 'a town' are equally irrelevant in signalling new events. The regular pattern language does allow phrases such as 'the attack against two villages' to be designated indefinite, as opposed to 'the attack against the village', which is marked as definite.

On the other hand, DPM is often tricked by modality – it cannot distinguish between events that actually happened and those that are merely discussed or referred to. This is one of the aspects of news texts that is dealt with in section 3.5.

Given the use of the primary and secondary activators, DPM is also necessarily domain-dependent. This is a common feature of shallow approaches to IE, and is therefore not surprising. As the method of construction of primary and secondary activators is not transparent, it is likely that the abstraction of activators for further domains would be a time-intensive operation.

MUCBRUCE

Cowie *et al*'s MUCBRUCE[8] system [Cowie *et al.* 92] performs an early statistical analysis of texts to decide which paragraphs can generate which types of incident frames (e.g. *arson*). Although this isn't really event recognition or distinction – especially as the paragraph/incident mappings are not unique, i.e. a particular paragraph can be marked as capable of generating more than one incident type – it could be seen as at least a suggestion of event structure within a text. For example, if the analysis suggested that paragraphs 1–3 of a document referred to an arson, whilst 4–6 described a kidnapping, then a multiple-event situation is an obvious assumption. In reality, however, the mechanism in MUCBRUCE is used primarily as a relevance filter for determining paragraph/incident type correlations.

2.4.3 Postsegmentation

Pre-analysis techniques, by definition, rely on a relatively surface-level form in order to derive a segmentation. It is not surprising, therefore, that attempts have been made to profit from the deeper (often semantic) levels of representation that result from linguistic analysis. Post-analysis techniques have been used in isolation, as an enhancement to frame merging, and in coordination with presegmentation.

NLTOOLSET

General Electric's TRUMPET [Rau & Jacobs 88], as suggested in section 2.4.2, works in conjunction with DPM as a top-down, domain-driven postsegmentation module operating on conceptual representations of sentences. Whilst it clearly duplicates some of the processing done by DPM, the designers claim that the interaction between the two modules aids overall system performance [Iwańska *et al.* 91].

TRUMPET uses a simple anaphora resolution algorithm, basic knowledge of causality, script-like predictive knowledge of stereotypical domain scenarios, and "*rudimentary temporal and spatial reasoning*" for determining event compatibility. Representations

[8] a subtle reference to a certain Monty Python song about philosophers

of events, called *E-Structs*[9], are merged unless their incident types, dates or locations are incompatible. TRUMPET performs well when it is presented with "*distinguishing conceptual information*" (i.e. explicit spatial or temporal shifts), but relies on DPM's analysis of activator definiteness for guidance in less overt situations.

CIRCUS

In one of its incarnations, Lehnert *et al*'s CIRCUS [Lehnert *et al.* 91] system employs a post-parsing segmentation approach as part of its discourse analysis component. Referred to as "*partitioning*", the process involves using textual cues to cluster task-specific semantic representations ("*consolidation structures* ") into partitions that weakly reflect text structure. Three classes of cue phrase are recognised: the first class introduce new events (e.g. *meanwhile*); a second class suggests generic (and usually irrelevant) events (e.g. *wave of*); and the third is used to signal separate events (e.g. *the day before*).[10] If a partition contains an event type that differs from that in a previous partition, domain- (and presumably incident-) dependent heuristics are used to recognise this as a boundary between multiple events.

INTERPRETEXT

Dahlgren *et al* [Dahlgren *et al.* 91] are one of the few groups to have attempted a formal linguistic analysis of discourse in ER – indeed one of the very few to have used any formal linguistics in IE.[11] The bulk of their discourse processing is based on Discourse Representation Theory [Kamp 81]. However, for event distinction their INTERPRETEXT system relies on a relatively simple model of segmentation. Texts are segmented after parsing, with event boundaries signalled by time changes, location changes, and certain cue phrases[12].

[9] not to be confused with the *E structures* described earlier in section 2.3, which are discourse structures

[10] It is interesting that the UMass team consider explicit temporal references such as these to be no more than members of a subclass of cue phrases. The referring potential such temporal phrases hold (and in a rich variety of forms) seems to us to warrant a more complex processing mechanism.

[11] Montgomery *et al* use a Government and Binding grammatical framework in their DBG IE system [Montgomery *et al.* 91] [Montgomery *et al.* 92], as did Robert Kuhns in his earlier News Analysis System [Kuhns 88].

[12] Including, apparently, "*in summary*", which is not an obvious indication of a new event.

2.4.4 Abductive frameworks

Abductive explanation [Hobbs *et al.* 90] is *inference to the best explanation*, and allows for a particularly subtle and elusive form of event recognition, as well as the elegant (if computationally expensive!) processing of various other discourse phenomena such as NP and pronominal anaphora resolution.

TACITUS

The relative success of TACITUS's implicit approach to event recognition is due to a combination of two theories, one implemented and the other manifested in news texts. The system performs a *minimisation* of predicate extensions, such that entities that share properties are assumed to be identical unless there is sufficient evidence to the contrary. This heuristic works because the texts that it is presented with usually conform to the Gricean maxim of relevance. Thus, by assuming that a text is coherent, it is fair to assume that the events and entities mentioned within it are related. These relations are made explicit as a result of the abductive inferencer's attempt to explain the truth values of sentences; minimisation then cements these relations by maximising the connections between sentences.

Due to the complexities involved in integrating temporal and spatial reasoning into the abductive mechanism, TACITUS actually delays time and location analysis until the template generation phase. TACITUS therefore exploits these reliable individuating criteria, supplementing the implicit event distinction performed under abduction with an explicit distinction process similar to that used by frame-merging systems.

2.4.5 Statistical methods

Although the most common role for statistical techniques in information extraction has typically been in the early text filtering stages [Lewis & Tong 92], some groups have used statistics for event recognition.

TTS

The Hughes Trainable Text Skimmer (TTS) [August & Dolan 92a], was designed with speed and domain portability in mind. It is perhaps unsurprising, therefore, that statistical techniques play an important part in the system.

Claimed to be analogous to discourse processing [Dolan *et al.* 91], the "topic grouping" performed by TTS builds a sentence-level image depicting the likelihood of individual sentences describing events of particular incident types. Figure 2.5 (adapted from [August & Dolan 92a]) shows the inputs and outputs of the topic grouping process, and illustrates how, through the use of a Gaussian mask, the "best fit" topic for each sentence is derived.

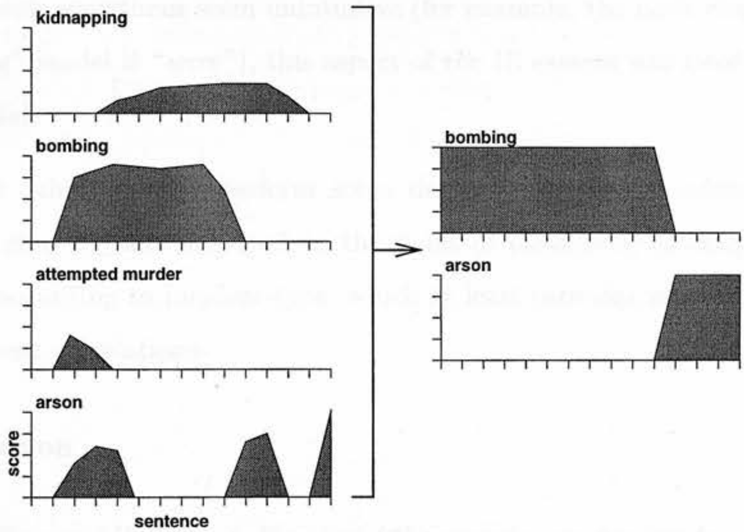


Figure 2.5: Topic grouping in Hughes' TTS

One of the advantages that Dolan *et al* claim with this approach is that generally high signals (with occasional momentary drops) are preferred over moderate continuous signals. This can be seen in the smoothing over the "arson" signal in figure 2.5. As is frequently the case with IE systems in the MUC domain, evaluation of individual system components is very hard to do. Dolan *et al* say that the small increase in overall performance for their IE system between MUC-3 and MUC-4 represents the arrival at a plateau constrained by the (statistical) approach they have adopted [August & Dolan 92b]. Whether this applies specifically to the topic grouping module or not is unclear, but it is at least suggested.

MUCBRUCE

The quasi-presegmentation performed by Cowie *et al*'s MUC-4 IE system [Cowie *et al.* 92] (see section 2.4.2) relies intensively on statistics. Although it does not truly distinguish between multiple events (except in the situation where events are of unique incident-types), it still warrants inclusion here as a statistical approach to the event recognition problem.

Word frequency models were derived for each incident-type from the 1300 text MUC-4 test corpus, with similar models constructed for detecting irrelevant text. Although the word models sometimes seem unintuitive (for example, the most common word in the "bombing" model is "*were*"), this aspect of the IE system was cited as one of the most successful.

Whilst many other systems perform some degree of statistical relevance filtering, MUCBRUCE goes beyond this level to the point of classifying (if only at the paragraph level) according to incident-type, which at least provides some indication as to paragraph-event correlations.

Lexical cohesion

Kozima and Furugori [Kozima & Furugori 93] propose a mechanism for measuring the semantic similarity between words in raw text. Using an English dictionary to automatically construct a semantic network, they use spreading activation as a means of computing lexical cohesion, a phenomenon that has long been recognised as useful for text structure analysis [Halliday & Hasan 76]. The cohesion relations between words are then used to build a "lexical cohesion profile" [Kozima 93] that is intended to show boundaries between cohesive segments in a text.

Their approach relies on the fact that word entries in dictionaries typically use other semantically related words in the definition. So, the word 'tree' would have 'plant' and 'treelike' in its definition. However, the degree to which semantically related terms are included in the definitions varies. For example, in the dictionary that Kozima and

Furugori use [LDOCE 87], the definition for 'tree' makes no reference to 'leaf'. This paucity of connotational and extensional terms injects an element of unpredictability into the mechanism. It is likely, however, that their approach is relatively robust to such discrepancies, so long as it is used at the sentence level and above.

As well as being computationally expensive, the usefulness of this method for distinguishing between multiple events in newspaper stories is unclear. One of the problems lies in the definition that Kozima and Furugori have of a text segment – "... *a sequence of clauses or sentences that display local coherence. It resembles a scene in a movie, which describes the same objects in the same situation.*" Evidently a theory that defines text segments as fragments that contain multiple references to the same objects (or, in this case, even "semantically related" objects) can be successfully implemented in a mechanism that is based on word frequency computations. As we shall see in section 9.9.2, this approach is not particularly successful in applications where text fragments can contain the same (and certainly semantically related) words without necessarily corresponding to the same event. Knowing that a document is lexically cohesive at the document-level (i.e. in Kozima and Furugori's terms exhibits strong semantic relations) is of no use if we are trying to determine how many actual events are described in that document.

TextTiling

Brief mention should also be made of Marti Hearst's TextTiling algorithm [Hearst 94], which uses domain-independent lexical frequency and distribution information to segment expository texts into contiguous, non-overlapping blocks. Although the ultimate goal of TextTiling is the rough analysis of segment content, impressive results have been obtained on identifying boundaries between segments in long texts. As with all statistical approaches, TextTiling is less useful when applied to shorter texts.

2.4.6 Text grammars

Given a body of texts that share a common domain, it seems plausible that these texts might also share certain material properties — such as the roles of the actors and objects in the discourse, and the nature of the actions and relationships with which they are associated. Vladimir Propp's analysis of Russian folktales in the 1920s [Propp 68] led him to conclude that, despite obvious surface differences between the stories, they nevertheless tended to possess the same structural properties in terms of both actors (*spheres of action* in Propp's terms) and actions (*narrative units*). Propp's original theory, which essentially constitutes a text grammar, has been built on by van Dijk [van Dijk 72] and others.

The rules of a text grammar embody expectations about the content and structure of a discourse, and would appear to be a useful knowledge source for reference resolution and the processing of unexpected input. The DBG message understanding system [Belvin *et al.* 89] uses a partially implemented text grammar in its MUC-3 instantiation as an extension to its otherwise straightforward frame merging approach. However, the team from Language Systems Inc. responsible for DBG found that the heterogeneous nature of the MUC documents was not well suited [Montgomery *et al.* 91] to the application of a text grammar. They claim greater success using text grammars for more homogeneous domains [Montgomery & Glover 84] such as the more formal military Space Event [Montgomery *et al.* 89] and Long Range Air [Stalls *et al.* 90] corpora. LSI's later versions of DBG omitted the text grammar, opting instead for explicit discourse markers [Montgomery *et al.* 92] (with an unrealised claim to be moving towards deictic centering [Rapaport *et al.* 89]) and, interestingly, speech acts such as *confirmation* and *repetition*, and highly application-oriented frame representations [Montgomery *et al.* 93].

2.4.7 Explicit approaches to event recognition and distinction

Having reviewed the various mechanisms that have been used to recognise events in documents, we can classify techniques according to whether they perform event recognition (ER) and distinction (ED) explicitly or implicitly. It is possible to find a system that is representative of each of the four permutations of this classification.

Explicit ER, explicit ED

The General Electric NLTOOLSET system is explicit both in ER and ED in that it not only looks for phrases that signal continuations of events (through the use of the "primary activators") but also phrases signalling event boundaries (via the "secondary activators").

Explicit ER, implicit ED

Most frame merging systems fall into this category. For example, BBN's PLUM explicitly seeks to merge events together (performing explicit ER). Frames that are not consolidated are therefore deemed to describe different events, so that event distinction is only handled implicit.

Implicit ER, explicit ED

There are very few systems that fall into this category. ITP's INTERPRETEXT postsegmentation approach constitutes one of them. It uses cue phrase, temporal and locative information to propose shifts in event.

Implicit ER, implicit ED

This is another rare category of approach. The only system representative of this type is SRI's TACITUS. As the bulk of the ER and ED is actually resolved through the use of the abductive interpretation system, with neither task being explicitly tackled, both can be said to be achieved in an implicit manner.

2.4.8 Recent developments

This literature survey has concentrated on IE systems developed in the context of MUC-3 and MUC-4. Because these evaluations made use of the same domain, developers were able to focus on improving existing technologies and exploring new ones, rather than on rapidly porting systems to new domains.[13] As a result, this body of work allows us to look at a broad range of techniques as developed over a period of years. Nevertheless, it is worth looking at how event recognition and distinction issues have been approached in the most recent MUC evaluations.

[13] This is not to say that no changes in the IE task were made between MUC-3 and MUC-4. One change was the increase in complexity of the template representation, rising from 18 slots to 24.

The most obvious innovations in MUC-5 [ARP94] were the adoption of new domains, *international joint ventures* and *electronic circuit manufacturing*; the use of a second language — documents from both domains were provided in English and Japanese; and a new representation formalism for the extracted information, which took the form of nested structures of objects. While this latter innovation makes it easier to represent multiple attributions in events, e.g. the fact that several people were involved in an event, and more importantly the representation of information relating to these people, at a higher level the relationship between templates and events still exists — though in terms of "event" objects instead of templates. It is not surprising, therefore, that systems fielded at MUC-5 exhibit a similar range of techniques as those described above.

The more recent MUC-6 [ARP96] evaluation continued the trend seen in MUC-5, i.e. the adoption of new domains as a means of encouraging domain-independent techniques, but was more structured in order that specific IE subtasks could be formally evaluated and to encourage research in areas seen by the organising committee as worthy of investigation. One of the areas deemed to be of interest was "deep natural language understanding", reflecting the generally acknowledged view that NLP components of IE systems were becoming increasingly shallow in approach. Because of this, the identification of noun phrase *coreference* within a document was isolated as a process meriting evaluation. Although the establishment of this task in MUC-6 serves to further underline the importance placed on discourse processing issues in IE, the task is of interest for another reason: the mechanism used to evaluate NP coreference decisions is relevant in evaluating event coreference. This is explored fully in section 8.2.4.

Systems fielded at MUC-6 show little difference in event recognition and distinction techniques. It is worth noting that the techniques used are often harder to identify in these systems than in those fielded at earlier evaluations, due to the nature of the domain and/or the narrative style used in the documents. In the MUC-6 domain (management changes), documents rarely contain references to similar (but distinct) management change events in addition to the core event. Instead, they typically contain commentary and relevant additional information such as stock price changes.

2.5 Summary

This chapter has presented an overview of previous research in fields relating to event recognition. We began by giving a brief introduction to information extraction, showing how its roots lie in the more naive ambitions of full text understanding, and the role that researchers such as Naomi Sager played in the rationalisation and focusing of the field into a more target-oriented application.

We also surveyed the more theoretical analyses that Iwańska, Moens and Dirk have carried out in the area of discourse structure and news texts, and cited evidence to suggest that news texts have quite idiosyncratic properties. We then presented an overview of the techniques that have been used in recent information extraction systems for the purpose of event recognition, and classified such approaches where possible. Each approach was then illustrated by one or more systems from the recent MUC evaluations.

As is suggested by our literature survey, there has been very little work on the evaluation of event recognition techniques. Only Hearst [Hearst 94] and Iwańska [Iwańska 91] give any serious consideration to issues of evaluation, and in the latter case (about which more in section 8.5.1) only qualitatively.

Having established the context in which we are working, the next chapter decribes the approach that we have taken to event recognition, and introduces the system that we have constructed to test our theories.

Chapter 3

System overview

3.1 Introduction

This chapter provides a general overview of both the theoretical and practical aspects of our approach to event recognition, and is organised into two main sections — **formalisms** for representing information at various levels of the event recognition process, and the **mechanisms** that are applied to these representations.

The main formalisms we introduce are those corresponding to CONTESS's input and, more importantly, to its output. In the context of the latter, we introduce the clause-event grid, and discuss some of the cognitive and mathematical properties that such grids possess.

Having presented the formalisms that we are using for event recognition, we will look at the general principles of the mechanisms involved. The overall architecture is discussed, and we introduce the language analysis modules and the event manager. This chapter is intended as a broad overview of the system as a whole — individual analysis modules, as well as the event manager, are presented in greater detail in subsequent chapters.

Formalisms

The input and output formalisms are represented diagrammatically in figure 3.1. The CONTESS system will be introduced later in this chapter in the section on mechanisms.

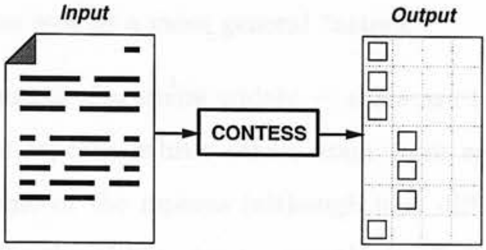


Figure 3.1: CONTESS input and output

Input

The input to CONTESS consists of a wide range of semantically linked American English documents which, though originating from multiple media sources in a variety of formats, have been compiled by professional analysts into text form. The following sections take a look at the nature of these documents, including the content and the physical and structural properties of the documents used.

3.2 Corpus and domain

CONTESS has been developed and evaluated using news reports about Latin American terrorist attacks.[1] In particular, we have been using the MUC-3/4 development and test corpus [Sundheim 91]. This constitutes 1600 texts — disseminated by the Foreign Broadcast Information Service (FBIS), an agency of the U.S. Government — that contain explicit references to both Latin American countries/nationalities (e.g. 'Bolivia/Bolivian') and an inflected form of a word associated with terrorist attacks (e.g. 'bombing'). The corpus is interesting from an information extraction viewpoint in terms of both its content and its form.

3.2.1 Corpus content

The documents in the corpus describe a wide range of terrorist incident types. This range is reflected in the set list of fillers for the "incident-type" field in the MUC-3/4 information extraction task: arson, bombings, kidnappings, hijackings, robberies,

[1] For a precise definition of what constitutes a Latin American terrorist attack for the MUC task, see [NRaD 92], section 1.0.

forced work stoppages, as well as a more general "attack".

The source of the documents also varies widely — some are from newspaper articles, television or radio news reports, whilst others stem from speeches, interviews and rebel communiques. Some of the reports (although it is difficult to tell how many) were originally compiled in Spanish, and subsequently translated before dissemination. We do not consider this to be a significant factor in this corpus.

3.2.2 Physical properties of documents

Each document consists of a header and a message body. The header, which is partially regularised, typically contains a unique document identifier code, a document date, a document location and a description of the media through which the information originated (e.g. 'communique', 'report'). Although all document headers contain identifiers, some lack date information, and a higher proportion have no explicit locative details. Figure 3.2 shows a typical MUC-3/4 document.

DEV-MUC3-0052 (NOSC)

BOGOTA, 21 JAN 89 (DPA) - [TEXT] IT HAS BEEN OFFICIALLY CONFIRMED THAT TWO LEADERS OF THE LEFTIST PATRIOTIC UNION [UP] WERE KILLED BY UNKNOWN PERSONS NEAR RIO BLANCO MUNICIPALITY, TOLIMA DEPARTMENT.

IT WAS LEARNED THAT CRISTOBAL PEREZ AND RUBEN DARIO RAMIREZ WERE SHOT TO DEATH BY A GROUP OF HEAVILY ARMED MEN.

SO FAR THIS YEAR, 12 MEMBERS OF THE UP, THE LARGEST LEFTIST POLITICAL FORCE IN COLOMBIA, HAVE BEEN KILLED BY RIGHTIST PARAMILITARY GROUPS. SINCE THE UP WAS FOUNDED IN 1985, APPROXIMATELY 900 OF ITS MEMBERS HAVE BEEN MURDERED.

ACCORDING TO THE ADMINISTRATIVE DEPARTMENT OF SECURITY (DAS), A RIGHTIST PARAMILITARY GROUP IS RESPONSIBLE FOR THE MASSACRE OF 12 PERSONS LAST WEDNESDAY. THE 12 WERE MEMBERS OF A LEGAL COMMISSION INVESTIGATING NEARLY 30 CASES OF MURDERS AND DISAPPEARANCES OF PEASANT LEADERS, SOME OF THEM UP MEMBERS, IN SEVERAL MUNICIPALITIES OF NORTHEASTERN COLOMBIA.

Figure 3.2: A typical MUC-3/4 document

Although paragraph boundaries are identifiable through the spacing conventions used in the corpus, these and other such boundaries (e.g. sentence boundaries) are marked up during the preprocessing stage of CONTESS (see section 3.5 below for details).

Furthermore, the vast majority (95%) of the corpus is in single case. This is to a certain degree unfortunate, as case is undeniably a useful feature in identifying sentence boundaries and proper names. FUNES [Coates-Stephens 92], a system for analysing proper names, relies heavily on mixed case; Coates-Stephens claims that the MUC-3/4 corpus is one of the few corpora to use single case, a claim that we would support. Reuters news agency, for example, use only mixed case in their newswires [2]. The effects of case on system performance are discussed in sections 9.8 and 9.10.

3.2.3 Structural properties of documents

The documents vary widely in terms of length. Average document length is approximately half a page of typed A4 paper (roughly 330 words), although the range in one subpart of the corpus (just 100 documents) was observed to be from 34 words up to 1074 words. Sentence lengths also vary across a large range — average sentence length is approximately 23 words, with minimum sentence lengths per document typically as low as 2 words (e.g. 'begin recording') and maximum sentence lengths approximately 200 words.[3]

Such variations in length are, however, of little importance in the processing that CONTESS does. The combination of splitting sentences into "clauses" during preprocessing and restricting parsing to be minimal and localised means that sentence length, traditionally a computational hurdle for text analysis systems, is not a problem.

Although the documents vary in terms of length, they often exhibit similar discourse structures. In section 2.3 we looked at some of the discourse phenomena that have been described in earlier research. Certain of these, notably the non-temporal ordering observed by Moens [Moens 92] and E-structures proposed by Iwańska [Iwańska 93], are readily identifiable in the corpus.

[2] Reuters Ltd., personal communication

[3] For the purpose of obtaining these statistics, sentences were defined as sequences of words delimited by periods, and words as sequences of characters delimited by whitespace.

3.2.4 Document relevance

The MUC-3/4 task involved extracting information about Latin American terrorist attacks from a corpus of documents deemed likely, through keyword search, to contain references to such incidents. More precisely, the task involved building a template and instantiating the slot-fills for each and every event in the document that qualified as a terrorist attack. This qualification is non-trivial to define — in fact, the definition of what counts as a "relevant incident" in the evaluation guidelines runs to approximately 1500 words [NRaD 92]. For example, incidents which are non-factual, non-recent, non-specific or contain military targets are deemed to be irrelevant.

As CONTESS is designed to be domain-independent, it has no proper concept of incident relevance. Indeed, one of the primary motives for exploring the multi-modular shallow approach take by CONTESS is to see whether the omission of such a concept is in fact feasible. As a consequence of this, it is expected that the segmentation performed by CONTESS will be at a finer level of granularity than that required by a system in the MUC-3/4 evaluations capable of distinguishing between relevant and irrelevant information. This has implications for the evaluation of our system, as we shall see in section 8.5.

It is worth pointing out, however, that the concept of relevance is employed to a small degree at various levels. In section 3.5.4 we shall see that, below the sentence level, reported speech is identified as being largely irrelevant to the event recognition process. Furthermore, experiments focussed on relevance at the level of entire documents reveal that improvements in segmentation performance may indeed be obtained through the use of such a filter. This will be discussed in section 7.4.

Output

We have chosen to represent the event structure of a document as a two dimensional grid. One dimension corresponds to the **clauses** in the text (ranging from 1 to n, where there are n clauses), whilst the other denotes **event numbers** (ranging from 1 to e, with e being the number of distinct events that the document describes).

3.3 Clause-event grids

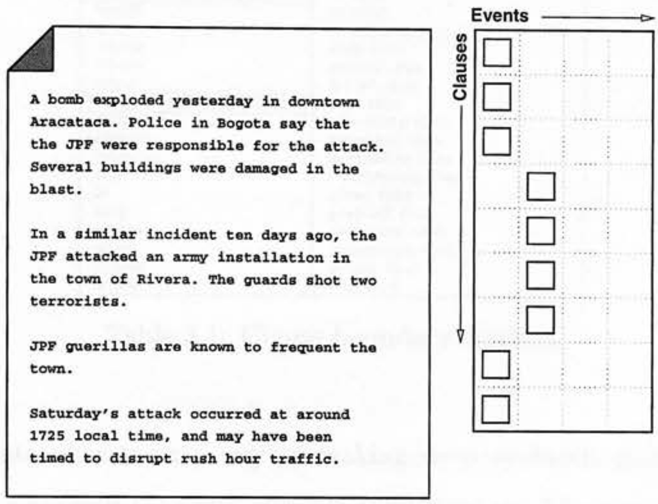


Figure 3.3: Sample text and corresponding event-clause grid

Figure 3.3 shows how the event structure of a particular sample text can be represented. In this example, a nine clause grid is deemed to consist of two separate events. Although it may be immediately clear from this figure, at least at some vague level, what we mean by clauses and events, it will be useful to present some definitions of the two primary components of our formalism.

3.3.1 Clauses

As the units of one of a clause-event grid's dimensions, "clauses" are easy to define. They correspond to fragments of text separated by various punctuation characters (essentially members of the class of "point" punctuation [Nunberg 90] below the sentence boundary marking level), relative pronouns, and both simple and compound conjunctive subordinators [Quirk *et al.* 72]. Table 3.1 shows the set of words and phrases that are considered as clause boundaries.

Hobbs *et al* use a similar strategy for segmenting sentences [Hobbs *et al.* 91] as part of their terminal substring parsing technique for TACITUS. Although we are concerned with the same body of texts, and the long sentences that such texts typically contain, our motivation is different to that of the SRI team. Instead of requiring that sentences

Punctuation	whenever
,	where
:	wherever
;	**Compound subordinators**
Relative pronouns	in that
which	so that
who	in order that
whose	such that
whom	except that
what	for all that
Simple subordinators	now that
although	providing that
though	provided that
because	supposing that
but	considering that
if	given that
how	granted that
however	assuming that
once	presuming that
unless	seeing that
when	and that

Table 3.1: Clause boundary markers

are segmented into clauses as a way of making deep syntactic parsing feasible, we simply require a finer unit than the sentence level gives us. The reason for this is that multiple events are occasionally described within the context of a single sentence. This is much less common within single clauses.

Given the simplistic approach to sentence segmentation that we have adopted, we clearly cannot claim any linguistic status for the clauses that result. Although such clauses are *usually* syntactically well-formed units, this is by no means universal. Their value stems primarily from the much finer granularity of description that they allow us in describing event structures.

The identification and marking up of clauses is discussed in section 3.5.

3.3.2 Events

As suggested in section 1.2, the concept of an event is very hard to define. However, as we are only interested in defining events with respect to information extraction, we shall restrict our definition to one that suffices for this task.

In the context of a clause-event grid, clauses can be seen as a *textual* dimension — that is, progression along the clause dimension represents discrete movement through the actual text, with intervals defined as those fragments of text delineated by clause boundary markers.

The *event* dimension, on the other hand, can be seen as a *physical* dimension — one that involves progression through either time or space. However, movement along this dimension is much harder to define than in the case of clauses. The reason for this is that whereas the distinguishing element for clauses is simply clause boundaries, the distinguishing element for events is less tangible. The MUC guidelines are quite clear in their definition of distinguishing properties for events — different fillers for either of the *location*, *date*, *category* or *perpetrator* slots indicate that a new template should be generated. In the context of the information extraction task, templates can be seen as analogous to events.

Definition 1 ***Event:*** *An occurrence that may be distinguished from other occurrences by generally understood (but rarely specified) means, for example temporal and/or spatial cohesion.*

We have adopted, therefore, an interpretation of the word *event* that is highly pragmatic and, in that it only really makes sense to talk about an event in the context of another event, an interpretation that is primarily contrastive.

A consequence of our definitions is that there must always exist a one-to-one directional relationship between clauses and events. That is, there must be no clause with multiple events, and there must be no clause which does not pertain to an event. We do not believe that these consequences are serious criticisms of our formalism, however. As we saw in section 2.3, single-multiple clause-event relationships are rare, and the usability of the formalism in such cases that do occur is not a problem.

The restriction that each clause pertain to an event is very rarely a problem. This may be a feature of the texts that we are concerned with, being as they are very informational in style, with little reflection or comment on incidents. This clearly has implications for system portability, and we discuss these issues (and those related to the above restriction) further in section 10.2.

3.4 Structure of grids

Now that we are a little clearer on the meanings of the two dimensions involved in clause-event grids, what meaning can we assign to the resulting grids? Using the text and corresponding grid in figure 3.3 as an illustration, we can see that the text describes two separate events, and that the two events identified correspond to clauses {1,2,3,7} and {4,5,6}.

Of course, it is important to note that there is no embedded structure implied here — we are concerned only with linear text segmentation, not text structuring[4] (a difference that can be viewed as one based on granularity). Neither is there any implicit event ordering suggested by the grid — the event corresponding to clauses {1,2,3,7} was only first *identified* before that contained in clauses {4,5,6}, and order of presentation of events in a news text is not always a reliable guide for order of occurrence of actual events.

Consequently, all we know from the grid in figure 3.3 is the following:

- the text it describes contains two events;
- those events are described by clauses {1,2,3,7} and {4,5,6}.

Further information may be inferred. For example, we can find out what the temporal order of events is by looking at the unified temporal analysis assigned to the constituent clauses of that event (see chapter 4 for details).

Furthermore, we might assume that the event described by clauses {1,2,3,7} is the more recent event, containing the "new information", and that the other event is providing some sort of context or background information. This is based on our knowledge of the structure of typical news texts [Noel 86], a knowledge source that has only rarely been exploited in IE tasks, and not something that we can simply infer from the grid as presented. There is, however, a clear correspondence between the grid in figure 3.3 and the E- (embedded-) structures of Iwańska [Iwańska 93].

In building grids (both manually and automatically), there is a very simple rule set that we adhere to. This can be stated as follows:

- begin in the top left hand corner of the grid;
- use new event columns, as necessary, from left to right.

[4] Marti Hearst's TextTiling algorithm (see section 2.4.5) is based on the same motivation.

These two rules restrict clause-event assignments within a grid to being either in the same column as a previous clause, or in the case where this would cause a conflict in terms of event relations, in the next available column to the right. Figure 3.4 shows all the possible 4-clause grids (we'll refer to grids with n clauses as n-clause grids). As

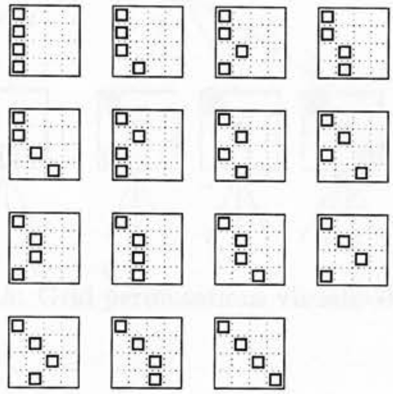


Figure 3.4: All possible 4-clause grids

we can see, there are fifteen possible 4-clause grids. It is useful to have some idea of how many possible n-clause grids there are for a range of values of n, if only so that we know what the chance is of randomly generating a particular grid. Table 3.2 shows, for low values of n, how many possible n-clause grids, $\mathcal{G}(n)$, there are.

n	1	2	3	4	5	6	7	8	9	10	...
$\mathcal{G}(n)$	1	2	5	15	52	203	877	4140	21147	115975	...

Table 3.2: Number of possible n-clause grids

The series is clearly monotonic increasing, divergent to infinity. Sloane and Plouffe [Sloane & Plouffe 95] [Sloane 94] refer to this series as the Bell sequence [Gould 71], and acknowledge it as being an important integer sequence. One way of visualising the sequence is as the number of nodes at descending levels in a tree, with a 1-clause grid at the root, as in figure 3.5. In the example shown here, only the first three levels are given, with shading used to clarify the relation between mothers and daughters in the tree. Extending the use of this terminology slightly further, we can see that there are two "families" at depth three of the tree — one family containing two daughters, and one containing three.

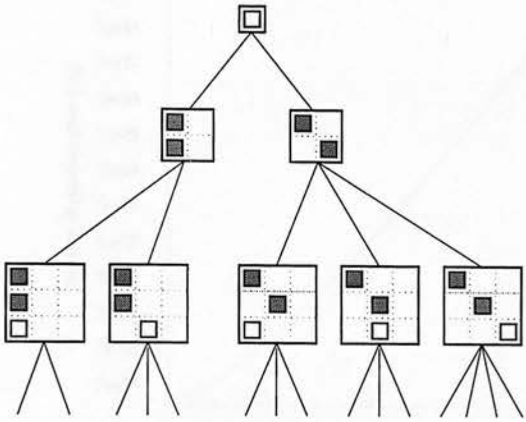


Figure 3.5: Grid permutations visualised as a tree

We can formally represent the relation between n and $\mathcal{G}(n)$ as an operation on a two-dimensional vector space, with one dimension corresponding to the number of clauses (i.e. depth in the tree), n, and the other denoting the number of families of specified sizes i_2^n. The equation is shown in 3.1, where $[\mathcal{F}(i,n)]$ denotes the number of families of size i at depth n, and $[\mathcal{F}(1,1)] = 1$ (i.e. there is only one possible 1-clause grid).

$$\mathcal{G}(n) = \sum_{i=2}^{n}([\mathcal{F}(i,n-1)] * (i-1)) + [\mathcal{F}(i-1,n-1)] + n \tag{3.1}$$

The graph in figure 3.6 shows the exponential growth in number of possible grids $\log(\mathcal{G}(n))$ plotted against length of grid n.

One of the consequences of this exponential growth in values of $\mathcal{G}(n)$ is that, as a grid description, it is highly unlikely that simply picking a hypothesis grid at random for realistic sizes of texts (for example, $n = 38$) will result in the correct grid being chosen. Although randomly proposing grids is a poor strategy for event recognition, it actually results in better scores than one might at first expect *for some classes of grids.* Section 8.2, and 8.2.3 in particular, present further mathematical aspects of clause-event grids in the context of evaluation.

Mechanisms

Having presented the various formalisms that we are using, we shall now proceed to

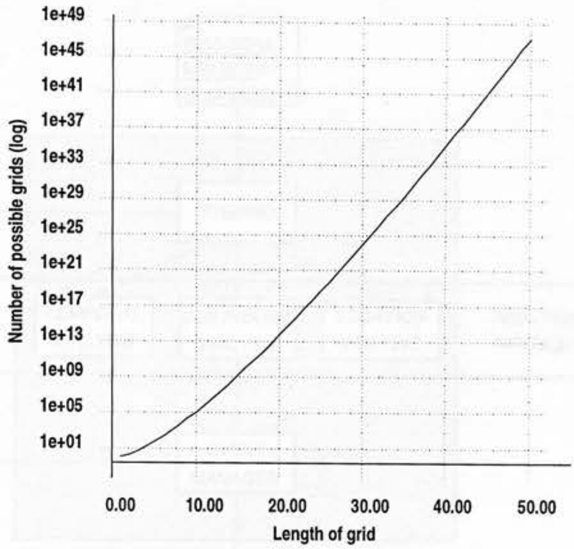


Figure 3.6: Exponential growth in grid possibilities with length

take a broad look at the mechanisms involved in our approach to event recognition.

As suggested in the introduction, our approach is a multi-modular one involving localised analysis of certain aspects of natural language. The open, modular design of CONTESS reflects our desire to create a system whose architecture permits the comparative evaluation of different types of linguistic knowledge. Figure 3.7 gives an overview of the architecture of CONTESS.

Processing begins with a MUC document being preprocessed (see below). A copy of the resulting marked-up document is then fed into each of the three analysis modules, which independently propose a set of constraints based on the language-specific analysis that they perform. These constraints are then passed to the event manager which, employing a suite of heuristics, constructs a segmentation for the original text.

Given that the analysis modules could theoretically be implemented so as to run in parallel, CONTESS's control flow can be seen as having three tiers — preprocessing, analysis and event management. Each of these facets of the segmentation mechanism will be introduced over the course of the remainder of this chapter.

The first aspect of this mechanism that we will look at is the preprocessing module. Although we do not consider preprocessing to be a major component of CONTESS, a certain amount of formatting and marking up needs to be done in order to convert

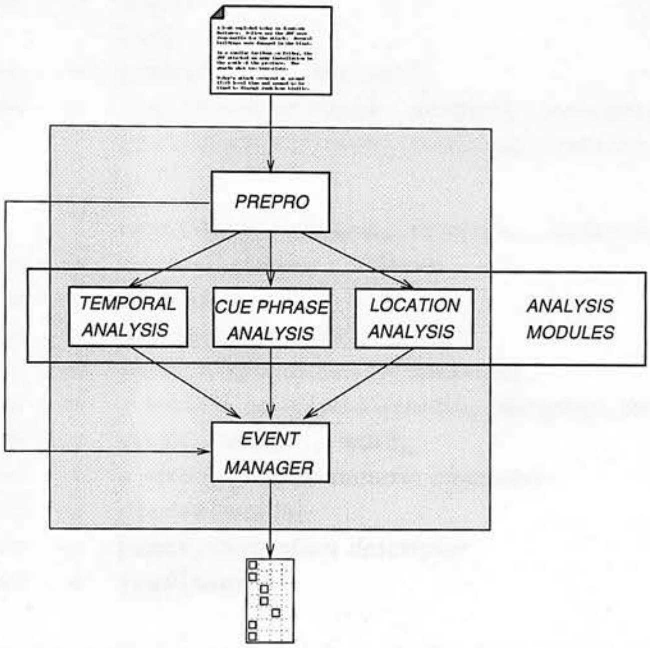


Figure 3.7: CONTESS system architecture

MUC documents into a form suitable for processing by Prolog programs. CONTESS is designed with integration into an IE system in mind, and whilst the task of preprocessing in the context of IE is one worthy of research, it is not one that we wish to address fully here. Consequently, we shall restrict our discussion of CONTESS's preprocessing module to a general overview of some of its more interesting properties.

3.5 Preprocessing

This module takes as input a raw MUC-3/4 document and performs a variety of normalisation and mark-up activities on the document. Each document is converted from its original free text form into a series of Prolog facts described by the following grammar.

document	$\rightarrow$	header
		body
header	$\rightarrow$	header(index, sentence).
body	$\rightarrow$	text(index, sentnum_1, newpara_1, sentence_1).
		text(index, sentnum_2, newpara_2, sentence_2).
		$\vdots$
		text(index, sentnum_n, newpara_n, sentence_n).
sentence	$\rightarrow$	clause_1, $\text{clause}_2 \ldots \text{clause}_n$
clause	$\rightarrow$	c[words]
index	$\rightarrow$	*document's unique identifier*
sentnum	$\rightarrow$	*position of sentence in document*
newpara	$\rightarrow$	*a boolean expression denoting paragraph boundaries*
words	$\rightarrow$	word_1, $\text{word}_2 \ldots \text{word}_n$
word	$\rightarrow$	*a string of alphanumeric characters*
word	$\rightarrow$	place#[words]
word	$\rightarrow$	punc#*punctuation descriptor*
word	$\rightarrow$	rsa#[words]

In the above, `rsa` stands for *reported speech analysis*, about which more in a moment. Figure 3.3 illustrates this grammar by showing an example text that has been preprocessed, highlighting a number of the specific activities performed.

The main preprocessing actions can be summarised as follows.

3.5.1 Headers

Each MUC-3/4 document has a unique identifier and a fairly regular header, which includes details of the document's temporal and spatial origins. Whilst some of the information, such as the source responsible for the document and its original medium and function, are not used by CONTESS, other header elements act to provide a reference for temporal and spatial expressions in the body of the text.

3.5.2 Boundaries

Word, clause, sentence and paragraph boundaries are identified through the use of various regular expressions. Words are defined as strings of alphanumeric characters delimited by whitespace or punctuation; clauses are marked up using the boundary markers listed in table 3.1; sentences are delimited by periods, exclamation marks, question marks and colons; and paragraph boundaries, being explicitly marked in the original documents using conventional means, are trivially mapped into the preprocessed form.

It should be noted that, due to effects of linearity and ordering in the preprocessing filters, some fragments escape clausal segmentation by virtue of having been previously marked up (for example, as being in a reported speech clause). However, we are not interested in the internal structure of such clauses and, therefore, this is by no means problematic.

before ...
TST1-DEMO-0001 BOGOTA, 4 FEB 90 (ACAN-EFE) – [TEXT]A BOMB EXPLODED YESTERDAY IN DOWNTOWN ARACATACA. POLICE IN BOGOTA SAY THAT THE JPF WERE RESPONSIBLE FOR THE ATTACK. SEVERAL BUILDINGS WERE DAMAGED IN THE BLAST. IN A SIMILAR INCIDENT TEN DAYS AGO, THE JPF ATTACKED AN ARMY INSTALLATION IN THE TOWN OF RIVERA. THE GUARDS SHOT TWO TERRORISTS. JPF GUERILLAS ARE KNOWN TO FREQUENT THE TOWN. SATURDAY'S ATTACK OCCURRED AT AROUND 1725 LOCAL TIME, AND MAY HAVE BEEN TIMED TO DISRUPT RUSH HOUR TRAFFIC.

... after

```
header(txt01, [c#[place#[bogota], punc#comma], c#[4, feb, 90, punc#openpar,
acan, efe, punc#closepar, punc#period]]).
text(txt01, 1, 1, [c#[a, bomb, exploded, yesterday, in, downtown,
place#[aracataca], punc#period]]).
text(txt01, 2, 0, [c#[rsa#[police, in, place#[bogota], say, that], the, jpf,
were, responsible, for, the, attack, punc#period]]).
text(txt01, 3, 0, [c#[several, buildings, were, damaged, in, the, blast,
punc#period]]).
text(txt01, 4, 1, [c#[in, a, similar, incident, ten, days, ago, punc#comma],
c#[the, jpf, attacked, an, army, installation, in, the, town, of,
place#[rivera], punc#period]]).
text(txt01, 5, 0, [c#[the, guards, shot, two, terrorists, punc#period]]).
text(txt01, 6, 1, [c#[jpf, guerillas, are, known, to, frequent, the, town,
punc#period]]).
text(txt01, 7, 1, [c#[saturday, s, attack, occurred, at, around, 1725,
local, time, punc#comma], c#[and, may, have, been, timed, to, disrupt, rush,
hour, traffic, punc#period]]).
```

Table 3.3: Preprocessing: before and after

3.5.3 Locations

Through the use of a restricted gazetteer of Latin American locations, place-name proper nouns are marked up and, where they occur adjacent to one another within the same clause, are incrementally grouped together. The example in table 3.3 contains a number of examples of single and multiple locative proper names that have been marked up. These have been underlined for illustrative purposes.

As we mentioned in section 3.2.2, the bulk of the MUC corpus is in single case. Clearly place-name identification is made harder by this fact, as Coates-Stephens points out in his doctoral thesis on the FUNES proper name analysis system [Coates-Stephens 92].

3.5.4 Reported speech

Earlier work by the author on IE techniques [Crowe 93] identified matrix clauses in reported speech (italicised in table 3.3) as being largely irrelevant for event recognition in particular — although often crucial for information extraction in general. Texts following the style of news reports typically make use of reported speech matrix clauses for providing elaborative information and commentary — rarely do we find that information crucial to the event recognition task is contained within such clauses.

The second sentence of the text in figure 3.3 contains an example of a reported speech matrix clause, i.e. 'Police in Bogota say that'. These clauses can be quite complex, for example 'A report published by the 'Cerigua' news agency – mouthpiece of the Guatemalan National Revolutionary Unity (URNG) – whose main offices are in Mexico, says that'. Clauses of this nature often contain temporal and locative references that incorrectly trigger the analysis modules, as we shall see in section 9.12.2.

Note, however, that such clauses also frequently include terms that are of great relevance in the IE task. The reported speech matrix clause 'The Guatemala army denied that', for example, contains an explicit *denial* of the information it is reporting. For this reason, we find it strange that Cowie *et al* take the step of simply stripping such prefatory clauses during the early stages of the IE process in their MUCBRUCE system [Cowie *et al.* 92].

Identification (as opposed to deletion) of these clauses is therefore useful for more than simply minimising false hits in event recognition; distinguishing the matrix clause from the reported event also facilitates the establishment of the *status* of reported events for IE systems, e.g. 'denied', 'confirmed' and 'accused'.

Reported speech analysis is performed using a series of regular expressions. Clause fragments that contain a reporting verb (e.g. 'say', 'added') and an instance of the word 'that' are marked up as reported speech. Having been marked up, they are then ignored by the analysis modules. As a result, the content of such matrix clauses cannot contribute to grid structure. Although there are cases where this approach fails (e.g. where there is no complementiser, or where a demonstrative is used and mistaken for a complementiser), it appears to be successful in over 95% of cases.

3.6 Analysis modules

The intention in this section is to give the reader a general idea of the motivation behind, and role of, the analysis modules in the overall context of CONTESS. Subsequent chapters will look more closely at the specifics of individual modules.

3.6.1 Motivation behind analysis modules

The analysis modules chosen for implementation within CONTESS reflect the six elements in table 3.4 that have been cited as useful cues for identifying event (or topic) shifts [Iwańska *et al.* 91], as stated in chapter 1.

• time	• tense
• location	• aspect
• topic	• cue phrases

Table 3.4: General event shift cues

The MUC-4 task documentation [NRaD 92] states that two partially instantiated message-templates within a document denote distinct events if their values differ for any of the four slots *perpetrator*, *location*, *time* and *incident type*.[5] This seems logical

[5] Examples of other slot types can be seen in appendix E, which contains a MUC template.

— any terrorist incident is a physical event (thus inheriting properties of time and location, given that they are typically described as occurring within the conventional model of the universe) of a particular kind (and thus classifiable into one of a set range of incident types) that does not occur naturally (therefore requiring the presence of a perpetrator).

We can see that there is a consensus both in general (in terms of the above six elements in table 3.4) and in particular (in the four-element list proposed in the MUC guidelines) that time, location and topic (or incident type), as the intersection of the two lists, are somehow fundamental in signalling event shifts. The *perpetrator* element, due to its domain dependent nature, does not occur in the general list of event shift cues.

There is an important distinction that can be made between those elements in the intersection and those that are exclusively members of the six-element list. Time, location and incident type are properties of *actual events*, and any change in either of them explicitly denotes a change in event. The other elements are simply linguistic strategies invoked, either through grammatical constraints or through stylistic concerns, to implicitly clarify or reinforce such changes. Of course, these linguistic strategies can and will extend to include references to time and location — indeed, a news report of a sudden physical incident (such as a terrorist report) that made no mention of the time or location of the event would seem strange. Such a text would, at the very least, be flouting the Gricean maxim of Relevance.

CONTESS's analysis modules have been selected with these points in mind, interpreting texts from the point of view of what we consider to be a promising subset of these various event shift cues. However, due to the early position we intend CONTESS to take up in an IE process, we have been unable to take tense and aspect into account in our analysis — the overhead in analysing these elements would have made the early processing position largely untenable. The MUC event shift cue of perpetrator also requires too much processing to warrant implementation within our constraints. On the other hand, cue phrases are relatively trivial to recognise — although, as we shall see in chapter 6, selecting the subset and the context of the cue phrases to use is somewhat harder.

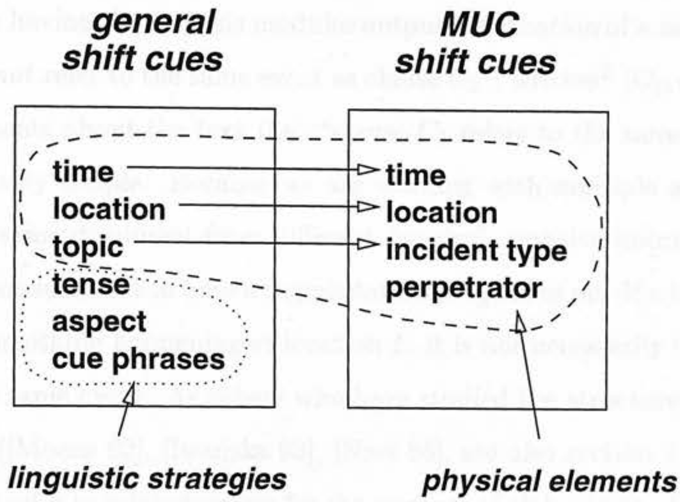


Figure 3.8: Physical elements and linguistic strategies in event shifts

The elements of time and location are much more complex than that of cue phrases, both to recognise and to interpret. However, as they are *relatively* restricted in their linguistic realisation, yet widely used in texts of the kind we are concerned with, they would seem to present a potentially rich source of event shift cues. One of the primary goals of our research is to see whether or not this is in fact the case.

3.6.2 Functionality of analysis modules

Despite being entirely independent of one another, two of the analysis modules — namely those focusing on time and location — share similar designs. Both modules undertake localised parsing within their own domains, and both build and interpret semantic constructs in the context of previously analysed fragments. They consequently share some of their key processing components, such as the parser and the grammar formalism interpreter.

The cue phrase analysis module, on the other hand, performs a much shallower level of analysis, and does not require a parser or interpreter. Instead, it uses pattern matching against a range of carefully selected phrases.

As we stated earlier on within this section on CONTESS's mechanisms, the output of each analysis module is a set of constraints describing, from a module-centric viewpoint, which clauses cannot refer to the same event.

The reasons for having the analysis modules output information of a negative form (i.e. "clause C_1 *cannot* refer to the same event as clause C_2", written[6] $\langle C_1, C_2 \rangle$) rather than positive statements about the text (i.e. "clause C_1 refers to the same event as clause C_2") are relatively simple. Because we are working with multiple analysis modules looking at the same document from different "angles", positive information from one angle alone is not sufficient to base a segmentation judgement on. If a text contains two references to something happening at location L, it is *not* necessarily the case that the two denote the same event. As others who have studied the structure of news reports have observed ([Moens 92], [Iwańska 93], [Noel 86], see also section 2.3), texts of this kind frequently refer to related events for the purpose of elaboration. Furthermore, the relation between such events are often precisely those physical elements in figure 3.8 that our analysis modules are searching for — i.e. time and location. In other words, the fact that events E_1 and E_2 both occurred at location L is likely to be the motivation behind their appearance in the same text.

Another problem with allowing analysis modules to provide positive information regarding clause-event coreference is that conflicts soon arise that are difficult to deal with in a satisfactory manner. For example, if the temporal analysis module were to propose that clauses C_1 and C_2 referred to the same event, whereas from the locative analysis module's viewpoint they were deemed to refer to different events, which module should we believe? One possibility would be to make use of confidence factors. However, it was found that a reliable and well-motivated method of assigning values that could be ported across modules was not feasible.

It is much more useful to have the analysis modules report negative information about a text. Thus, if one of the analysis modules proposes a constraint requiring C_1 and C_2 to refer to different events, this information is useful regardless of any constraints proposed by other analysis modules. CONTESS currently has no mechanism for dealing with degrees of constraint — a constraint either exists or does not exist, so that identical constraints proposed by multiple analysis modules have no greater effect than just one constraint proposed by a single module.

[6] Clauses are actually defined using sentence number and clause number. This representation is simplified here for purposes of illustration.

3.7 Constraint-based event manager

Because the task of the analysis modules is restricted to ruling out certain elements of a text segmentation, the event manager receives a very sketchy image of what the final segmentation grid looks like. The output from the analysis modules determines only what aspects of the grid cannot look like. The behaviour of the event manager therefore has a highly significant effect on the segmentation grids produced and, consequently, the performance of CONTESS as a whole.

The constraints proposed by the analysis modules are input to the event manager, whose role it is to build a grid describing the event segmentation of the text. To do this, it uses the constraints in conjunction with a range of heuristics that operate at various levels of the document — at the text level, the sentence level and the word level. Because the heuristics are based on various surface aspects of the text, the event manager must also have access to the original document (in preprocessed form), as shown in figure 3.7.

Despite the use of the constraints and the heuristics, the event manager is still often faced with more than one possible grid outcome. To resolve this situation, the event manager has access to a set of clustering strategies. These strategies allow the manager to fully and unambiguously specify the precise form of the segmentation grid.

The clustering strategies, heuristics and the event manager in general will all be examined in much greater detail in chapter 8.

3.8 Summary

This section has introduced the main formalisms and mechanisms employed by CONTESS in its approach to event recognition. Some structural properties of the corpus, the input formalism, were discussed, as well as the concept of document relevance with respect to CONTESS. The output formalism consisted of the clause-event grid, which was described in terms of its mathematical properties, including the relation between grid length n and possible grid forms $\mathcal{G}(n)$.

The three main layers comprising the mechanisms in CONTESS were also introduced.

We showed some of the preprocessing activities carried out, including reported speech analysis and location lookup. In terms of analysis, we discussed the motivation behind, and functionality of, the three analysis modules. Finally, we presented the Event Manager, which uses a combination of the constraints output by the analysis modules and various heuristics to propose grid structures.

The following chapters examine in detail the role of the three analysis modules and the event manager.

Chapter 4

Temporal analysis

4.1 Introduction

Whether it be due to Grice's maxim of Relevance, journalistic style, or some other unwritten rule of story telling, the temporal positioning of a reported incident is fundamental to its narration. This chapter examines the role that temporal phrases play in news reports, and describes the Temporal Analysis Module (TAM) in detail.

We begin by taking a close look at the linguistic techniques through which a reader may be provided with temporal information. We continue by looking at one particular source of information, a constituent we call a *temporal phrase*, and describe our method of identifying and classifiying such phrases.

Having introduced the linguistic feature that we are concerned with, the chapter then addresses issues of implementation. In particular, we propose a technique for extracting temporal information from a text through the recognition and analysis of temporal phrases. Sections 4.5 and 4.6 describe the parsing and interpretation strategies adopted within the TAM. Section 4.7 follows the route of a sample document through the TAM, showing at each stage the level of representation used.

4.2 Temporal information in discourse

It has long been acknowledged (see for example [Hirschman & Story 81]) that there are multiple sources of temporal information in narrative texts. These sources can be classified as follows.

- **Explicit** — includes phrases such as 'on Wednesday', 'tomorrow' and 'in June', as well as non-phrasal elements such as tense. As we shall see, there are in fact fewer examples of completely explicit temporal information sources in narrative texts than one might initially think.

- **Implicit** — the convention of narrative time progression, or iconic sequencing [Fleischman 90], dictates that, unless otherwise explicitly marked, the order of events in a text reflects the order of occurrence of events.

- **Anaphoric** — this source denotes temporal information derived from the interpretation of an expression as referring to a previously mentioned event. Examples of temporal anaphora are 'then', 'that year', 'three days later' and 'afterwards' [Hirst 81]. Section 4.6.3 in this chapter presents our approach to anaphoric temporal expressions.

CONTESS is concerned with all three sources of temporal information, though to varying degrees and at different stages of processing. A wide range of explicit and anaphoric temporal cues are identified and interpreted during temporal analysis. Implicit temporal information, however, is only (and appropriately) *indirectly* processed within the Event Manager (see section 7.3).

There are several reasons for this. Moens [Moens 92] has shown that the reliability of implicit temporal sources depends on the discourse type of the document in question. The sequential correspondence between events in the text and those in the real world can be seen as a defining feature of narrative texts — Labov [Labov 72] describes such texts as "a method of recapitulating past experience by matching [...] clauses to the sequence of events (it is inferred) actually took place".

Although the news reports that we are concerned with can be generally described as narrative texts, they are in fact a specialised subclass of this discourse type. As we stated in section 2.3, the order of events in news reports rarely corresponds chronologically to the order in the world being described. Consequently, we cannot fully rely on iconic sequencing to provide us *directly* with an implicit source of temporal ordering information.

4.3 Temporal phrases

4.3.1 Definition

The temporal phrases that we are concerned with have previously been classified, at least informally, as *time relators* ([Quirk *et al.* 72], section 19.35) and *overt time phrases*. The following examples are illustrative of our concept of temporal phrases. The sentences are taken from the MUC-4 development corpus; temporal phrases are highlighted in italics.

- it has been officially reported in la paz that a bomb attack occurred *early this morning* somewhere near the place where bolivian president jaime paz zamora and peruvian president garcia are to meet today.
- 3 members stationed in la union department reported that two fmln rebels were killed and five others wounded during clashes near el refugio farm, san miguel jurisdiction *in the afternoon of 27 november*.
- leftist guerrillas *early tuesday* attacked a residential neighborhood in northern san salvador where many government and military leaders have their homes.
- the murder took place *at 2220 last night* and, *30 minutes later*, the terrorists left messages with local newspapers claiming to be the assailants.

Given that the above fragments are representative of what we consider to be temporal phrases, how can we describe their defining characteristics? We have defined temporal phrases as below.

Definition 2 ***Temporal phrases:*** *overt constituents that, although syntactically optional, play a crucial role in semantically linking punctual discourse events to points in time.*

There are a number of important issues in this definition. We restrict the definition of temporal phrases to *overt* constituents in order to rule out elliptical temporal information sources (such as tense and aspect and iconic sequencing). Furthermore, we say that they are overt rather than explicit, as we do not want to exclude anaphoric temporal expressions (as described in section 4.2 above).

Temporal phrases such as those in the example above can be omitted with little or no effect on the grammaticality or sense of the original sentence. Clearly, there are a small number of verbal constructions (such as 'happened', or 'took place' in the example above) that rarely occur without temporal phrases. The sentence containing the fragment 'the murder took place, and the terrorists left messages ... ' is syntactically well-formed and, although it might seem a little strange, it is not hard to envisage contexts where this sentence would make perfect sense (e.g. 'contrary to previous reports, the murder took place, and ... ').

In fact, verbal constructions such as 'took place' are useful heuristics for determining whether or not a phrase falls within our definition of a temporal phrase. Thus, we have the following grammaticality judgements:

$$\text{X took place} \left\{ \begin{array}{l} \text{at 0700} \\ \text{yesterday morning} \\ \text{on the afternoon of 2 july} \\ \text{2 days ago} \\ \text{last night at about 2115} \\ \text{until march 3rd*} \\ \text{since friday*} \end{array} \right\}$$

Clauses modified by phrases beginning 'until' and 'since' (and others like 'while' and 'once') must be *durative* ([Quirk *et al.* 72], section 15.27). As we stated in section 1.2, Quirk *et al.*'s situation typology classifies the events that CONTESS is concerned with as *punctual* situations — that is, momentary (and, to a lesser extent, transitional) events and acts. The final part of our definition of temporal phrases therefore reflects this constraint.

4.3.2 Identification

Construction of the temporal analysis module began with the compilation of a list of temporal phrases. This was achieved through an iterative process, which began with the author's intuition and continued through the use of Quirk *et al.* [Quirk *et al.* 72] and a thorough inspection of the 430,000 word development corpus, resulting in a list of approximately 30 Perl regular expression forms [Wall & Schwartz 91] as shown in table 4.1.[1]

`\bmorning[s]?\b`	`\bmonth\b`	`\bnov[\be]`
`\bnoon\b`	`\bweek\b`	`\bdec[\be]`
`\bevening[s]?\b`	`\bjan[\bu]`	`\bgmt\b`
`day[s]?\b`	`\bfeb[\br]`	`\b[012]\d\d\d\b`
`\bnight[s]?\b`	`\bmar[\bc]`	`\bdate[\bs]`
`\blast\b`	`\bapr[\bi]`	`\bstill\b`
`\bnext\b`	`\bmay\b`	`\bnow\b`
`\bearl`	`\bjun[\be]`	`\bsoon\b`
`\blate`	`\bjul[\by]`	
`\btomorrow\b`	`\baug[\bu]`	
`\bhour[\bs]`	`\bsep[\bt]`	
`\byear[\bs]`	`\boct[\bo]`	

Table 4.1: Regular expressions for identifying temporal phrases

The development corpus was then examined for instances of each of the regular expressions, and phrases were manually extracted from the matching fragments of text. As the list of regular expressions was designed to match as many temporal phrases as possible, significant manual editing of the resulting fragments was necessary — for example, in the case of homonyms such as 'may', as in 'individuals who *may* have seen the holdup men at close range'. Table 4.2 shows a subset of the temporal phrases identified in this manner along with the relevant keyword, and is intended to give a flavour of the kinds of phrases identified.

morning		*evening*	
	at 0600 this morning		on that same evening
	on the morning of 3 november		the evening of 1 november
	yesterday morning at 0730		friday evening
	...		...
noon		*day*	
	this afternoon		today
	today at noon		2 days ago
	at noon of 5 november		early tuesday
	...		...

Table 4.2: Identification of temporal phrases

The following sections look in greater detail at the inner workings of CONTESS's Temporal Analysis Module. We begin by giving a quick overview of the information flow

[1] In the regular expression formalism used, `\b` denotes a word boundary, `?` an optional character, and `[...]` a set of alternative characters. Full details can be found in [Wall & Schwartz 91].

and processes involved, and then proceed to describe the components that make up the Module and the relationships between them.

4.4 Overview of Temporal Analysis Module

The architecture of the Temporal Analysis Module (TAM) is shown in figure 4.1. The input to the TAM consists of a single preprocessed document, and the output is a set of constraints stating which clauses, in the opinion of the TAM, cannot refer to the same event (see sections 3.5 and 3.6.2 in the previous chapter). The controlling mechanism of the TAM is the *sentence processor*, which reads in a sentence at a time and then passes each clause to the rest of the module for analysis.

The bulk of the TAM consists of the *island-driven parser* and the *semantic interpreter*. The parser is alerted to possible temporal phrases through the presence of *keywords*, which constitute "islands of certainty" for the parser to build outwards from.

The parser outputs bundles of feature structures to the semantic interpreter which, primarily in conjunction with a *knowledge base*, interprets them to provide explicit temporal representations.

The interpreted temporal information is then passed back into the sentence processor, where it becomes available as a referent for subsequent temporal anaphora.

Finally, once the document has been fully processed, the sentence processor examines the resulting temporal representations and produces a set of constraints describing the relationships between them.

4.5 Syntactic analysis of temporal phrases

This section describes in greater detail the operation of the syntactic analysis components of the TAM: keywords, the island driven parser, and the grammar of temporal phrases.

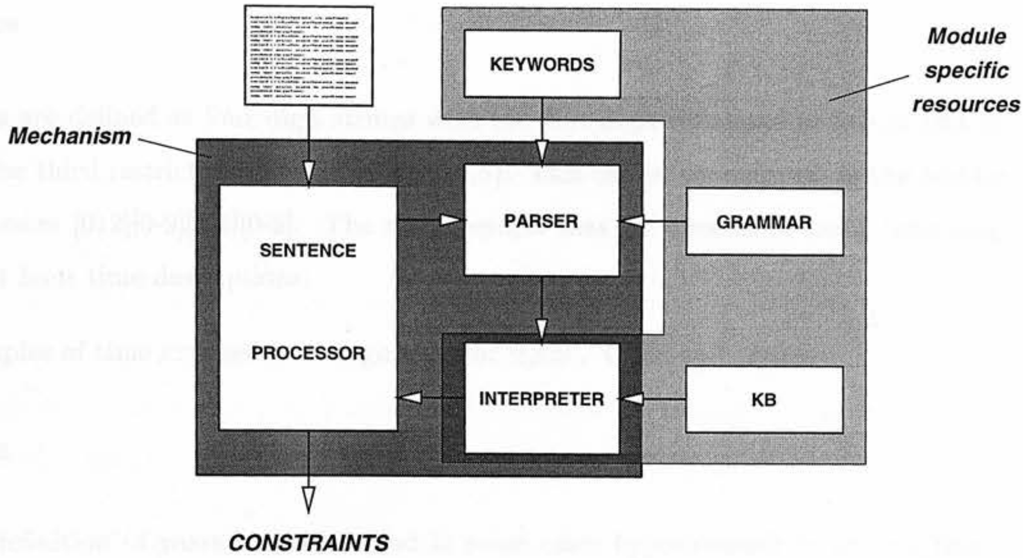


Figure 4.1: Temporal Analysis Module (TAM) architecture

4.5.1 Keywords

There are 56 explicit keyword forms (including morphological variations) plus two generic keyword types, times and years.

Explicit keywords

The keywords were derived through the manual analysis of news texts, with further terms being added as a result of the expansion of classes such as months, days and day parts. Explicit keywords are shown in table 4.3.

> morning, afternoon, evening, night, nights, monday, tuesday, wednesday, thursday, friday, saturday, sunday, january, february, march, april, may, june, july, august, september, october, november, december, jan, feb, mar, apr, jun, jul, aug, sep, oct, nov, dec, hour, hours, day, days, week, weeks, weekend, month, months, year, years, today, tonight, tomorrow, yesterday, daybreak, nightfall, dusk, midnight, dawn, gmt

Table 4.3: Temporal analysis module keywords

Times

Times are defined as four digit strings with the first digit restricted to one of {0,1,2} and the third restricted to one of {0,1,2,3,4,5}. This can be represented as the regular expression [012][0-9][0-5][0-9]. The news reports that we have been using invariably use 24 hour time descriptions.

Examples of time expressions recognised are: '1259', '0730' and '2400'.

Years

The definition of years is similar (and in some cases hyponymous) to that of times. Years are either two or four digit strings (to cover both YY and CCYY formats), and can be defined in regular expression form as [789][0-9] and [12][90][0-9][0-9].

Restricting year expressions to a subset of numeric strings means that, in fragments such as the example below, non-year numeric forms are prevented from acting as years.

- on 18 december 18 of the prisoners...

Examples of year expressions include: '89', '1990' and '1936'. It was observed that references to years in YY format were chronologically restricted, so that although the relatively recent '1987' can be found as simply '87', the more distant '1942' is rarely, if ever, shortened to '42'.

Ambiguity in times and years

Clearly there is some overlap between time and year expressions. For example, '1942' can be a 24 hour time expression as well as a year. However, context (in particular, the use of prepositions 'in' and 'at') typically resolves ambiguity. Hypothetical cases such as 'the 1955 murder of drugs baron Sr. X' would be interpreted by CONTESS as taking place in the mid-fifties, rather than after dinner, simply because of ordering in the keyword list. Were the TAM to encounter a phrase such as 'the 1955 express to Bogota', on the other hand, our strategy would need modifying.

4.5.2 Islands

Having identified a keyword, the parser proceeds by building up a word string bidirectionally and then attempting to parse it. Words are added to the string if they are known (i.e. in the lexicon); integers are assumed to be known.

If the parser (a simple top-down, left-right PATR [Shieber *et al.* 83] interpreter) can parse the string, then it produces an uninterpreted semantic representation of the phrase. If, however, the parser fails, then the input string is incrementally reduced until either parsing succeeds, or there are no words left in the string to parse.

The string deconstruction procedure simply uses Prolog's backtracking mechanism to recursively remove words from the string until only the keyword is left. Given a string with five words, where W_0 is the keyword and the string extends 2 places to the left to word W_{-2} and 2 places to the right to word W_2, the resulting deconstruction pattern is shown in table 4.4.

step	**string**				
1	W_{-2}	W_{-1}	W_0	W_1	W_2
2	W_{-2}	W_{-1}	W_0	W_1	
3	W_{-2}	W_{-1}	W_0		
4		W_{-1}	W_0	W_1	W_2
5		W_{-1}	W_0	W_1	
6		W_{-1}	W_0		
7			W_0	W_1	W_2
8			W_0	W_1	
9			W_0		

Table 4.4: String deconstruction mechanism in island-driven parsing

The first string that can be parsed successfully is used. If the parser cannot parse the smallest possible string, i.e. the keyword alone (an example of a keyword that is also a temporal phrase is 'tomorrow'), then the parser abandons that keyword and moves on to the next.

Once a string has been successfully processed, the parser moves to the end of that string and continues looking for keywords. This avoids potential problem where a string containing multiple keywords is interpreted more than once.

4.5.3 A grammar of temporal phrases

The parser has access to a PATR formalism unification-based grammar. The grammar contains 31 phrasal rules and 124 lexical entries, and has been modified to allow rules to include regular expression-style forms such as alternatives and optional elements.

The language of temporal phrases can be seen as a sublanguage and, as is typically the case in the construction of grammars for restricted domains, grammar design was initiated by manual analysis of the texts. Using such techniques as keyword search (with, for example, some of the keywords in section 4.5.1), instances of temporal phrases were extracted, collated and analysed.

To illustrate, we will show how a particular collection of phrases is represented in the grammar and lexicon, and present the rules and lexical entries that are required to parse these phrases. The phrases that we shall consider include:

- early this morning
- early tuesday
- yesterday morning at 0730
- late the following year
- on friday evening at 1903
- ...

We found that there exists a set of phrases following the pattern (MOD) X (AT_TIME), where MOD is an optional modifier such as 'early' or 'late', and AT_TIME is an optional time expression such as 'at 1903'. A variety of constructions were observed as occurring in the position of X, such as 'this evening', 'last thursday', 'that same night' and 'yesterday morning'. Each of these four examples is an instance of a separate type of phrase in the grammar and, as a result, we wanted to be able to stipulate that any of these phrasal categories could appear in this top level rule.[2]

The rule describing the above construction makes use of regular expression extensions that we have included in our grammar formalism. These consist of:

[2] These phrases are also top-level temporal phrases independently.

- opt(element) — specifies that element is optional
- set([element$_1$, element$_2$, ...element$_n$]) — any of the alternatives element$_1$ through element$_n$ may be used
- unisem(rule, [element$_1$, element$_2$, ...element$_n$]) — recursively unifies the features of the elements in the list with the features of the rule

The rule therefore looks as follows:

```
%Grammar rule R0
R0 ---> opt(A),set([B,C,D,E]),opt(F):-
        R0:syn:cat <=> s,
        A:syn:cat <=> premod,
        B:syn:cat <=> s,
        B:rule <=> r1,
        C:syn:cat <=> s,
        C:rule <=> r2,
        D:syn:cat <=> s,
        D:rule <=> r2i,
        E:syn:cat <=> s,
        E:rule <=> r3,
        F:syn:cat <=> attime,
        R0:rule <=> r0,
        unisem(R0,[A,B,C,D,E,F]).
```

The rule format given above is typical for a temporal phrase — top-level rules have a rule feature that defines the rule name. A consequence of the regular expression extensions is the need for the unisem/2 Prolog clause, which recursively performs feature structure unification between the two sides of the rule.

The following lexical entry represents an example of the optional premodifier (syntactic category (syn:cat) premod) in the above rule.

```
X ---> [early]:-
        X:syn:cat <=> premod,
        X:sem:mod <=> early.
...
```

These entries are defined as having the premod syntactic category (the value of the syn:cat feature path), and a semantic feature mod (modifier) with appropriate values.

One of the alternative elements in our example top-level rule (designated by the variable C, which corresponds to the rule with the unique identifier r2) is for phrases such as 'last thursday'. To further illustrate the grammar, we shall follow the expansion of this part of the rule down to the level of lexical entries.

If we look at rule r2, we can see that it refers to four further entries in the grammar, one of which is optional and two of which are interchangeable.

```
%Grammar rule R2
R2 ---> A,set([B,C]),opt(D):-
        R2:syn:cat <=> s,
        A:syn:cat <=> det,
        B:syn:cat <=> unit,
        C:syn:cat <=> day,
        D:syn:cat <=> daypart,
        R2:rule <=> r2,
        unisem(R2,[A,B,C,D]).
```

Although in this context we are not using this rule as a top-level rule, it can itself function as such — as can be seen by the presence of the distinguished symbol s as the syntactic category of the left hand side of the rule.

An example of the obligatory element required by this rule is the definition for the lexeme 'last' (as in the above example). This is defined in the lexicon primarily in terms of a connective (semantic feature conn), an amount (feature amnt) and a referent to which the phrase is relative (feature rel, see section 4.6.3 below for further discussion of relative temporal phrases).

```
X ---> [last]:-
        X:syn:cat <=> det,
        X:sem:conn <=> '<',
        X:sem:amnt <=> 1,
        X:sem:rel <=> doc.
...
```

The next element of the phrase has two possible alternatives. The first, designated as having the category unit, represents units of time such as minutes, days and weeks. An instance of a unit from the lexicon is shown below.

```
X ---> [week]:-
```

```
        X:syn:cat <=> unit,
        X:sem:unit <=> week.
...
```

The grammar rule for r2 stipulates that, instead of units of time, a category defined as day can occur in its place. This category includes days of the week, as shown in the following example.

```
X ---> [friday]:-
        X:syn:cat <=> day,
        X:sem:unit <=> week,
        X:sem:day <=> fri.
...
```

Note that 'friday', as with the definition for 'week' above, has a semantic unit feature; furthermore, they both have the same value for this feature. This is required to correctly interpret statements such as

- the president will leave peru next friday.

where one interpretation of 'next Friday' is 'the Friday of next week'.[3]

The final element of rule r2 is the syntactic category daypart, which corresponds to parts of day such as morning, afternoon, evening and night. Although all of these terms have, to varying degrees, somewhat fuzzy definitions, it is not particularly important to resolve them to precise times — it is only necessary to know that, at least within the same text, they refer to different times.

```
X ---> [morning]:-
        X:syn:cat <=> daypart,
        X:sem:daypart <=> morning.
...
```

Informal research resulted in the classification in table 4.5 being generally agreed on as a good set of working definitions for day part descriptions. However, until the

[3] There is definite disagreement amongst native (British English) speakers as to which Friday is being referred to in these constructions.

European Commission introduces the relevant legislation, we feel that such dubious classifications are best avoided.[4]

0000–1159	morning
1200–1745	afternoon
1746–2059	evening
2100–2359	night

Table 4.5: A possible classification of day part descriptions

It is possible, on the other hand, to make use of *part* of this classification. As we do not want temporal expressions such as '0900' to be interpreted as anything other than morning, we automatically add a daypart feature with value morning to any expressions in 24-hour format between midnight and 11:59, and the value night to expressions from 21:00 to 23:59. Remaining boundaries are, we feel, too vague to warrant similar treatment.

As we have now fully illustrated all rules and lexical entries relating to the rule r2, there only remains the optional category attime from our top-level rule. This describes expressions such as 'at about 1230 GMT'.

```
X ---> A,opt(B),C,opt(D):-
        X:syn:cat <=> attime,
        A:syn:cat <=> prep,
        B:syn:cat <=> about_word,
        C:syn:cat <=> time,
        D:syn:cat <=> timezone,
        unisem(X,[A,C,D]).
```

The four elements that this rule contains are extremely simple. The first, which is obligatory, includes prepositions such as 'at' and 'after'.

```
X ---> [after]:-
        X:syn:cat <=> prep,
        X:sem:when <=> after.
...
```

This element is obligatory in order to prevent the parser from interpreting isolated numeric expressions such as '1500' as times — the minimum context that such expressions

[4] It has been pointed out that, according to this classification, 10.30pm is technically the middle of the night.

require is a preposition.

The optional category about_word simply functions as syntactic sugar. No semantic information is contained within it.

```
X ---> [about]:-
        X:syn:cat <=> about_word.
...
```

The second obligatory element in the attime category is the numeric temporal expression itself. In the rule shown below, poss_time/1 checks that the numeric element is a legal temporal expression (as described in section 4.5.1), and find_daypart/2

```
X ---> [Y]:-
        X:syn:cat <=> time,
        X:sem:hour <=> Y,
        poss_time(Y),
        find_daypart(D,Y),
        X:sem:daypart <=> D.
```

The final element, again optional, is a time zone specifier such as gmt and edt. The contribution this makes to distinguishing between temporal expressions is somewhat dubious — one could imagine a document that referred to both '1230 gmt' and '1230 edt', but this would be highly unusual.

4.6 Semantic analysis of temporal phrases

The parsing component of the TAM produces a certain amount of semantic information, some of which we have hinted at in the context of the previous section. This section will look in more detail at the semantic representations that the TAM uses, and will describe the typology used in these representations. It will also examine the role of knowledge sources for temporal analysis, describe the semantic interpretation that the TAM performs, and look at the problem of processing relative temporal phrases.

4.6.1 Typology of temporal relations

The semantic typology used within the TAM has thirteen major features, seven of which have a fixed set of possible values (approximately four each on average), and six of which have (finite, but broad ranging) numeric values such as hours and dates. The full set of features and values is presented in table 4.6.

`rel`

In section 4.2 we stated that temporal expressions are commonly anaphoric. As we shall see in section 4.6.3 below, all temporal expressions can be interpreted as being anaphoric to varying degrees. One broad but important distinction we can make between temporal expressions relates to the referent to which the expression refers. All temporal expressions are interpreted as being relative to either the document date or the date of an event in the text, and the `rel` feature dictates which of these cases holds for a particular expression.

`conn`

If a temporal expression is to be relative to a previously mentioned time, then it must be chronologically either earlier, later, or the same. The `conn` feature denotes the direction in which the relation holds.

`amnt`

Continuing in the theme of relative temporal expressions, we also need to stipulate how far in a particular direction our expression differs from the referent. The numeric part of this information is stored in the `amnt` feature. Although the syntactic component of the TAM can interpret expressions such as that in 'a few days ago', qualitative phrases such as 'a few' are translated into figures so that the expressions in which they appear can be included in the quantitative reasoning process that the semantic component of the TAM performs.

`unit`

The final part of the information required to resolve temporal anaphora is the `units` in which to measure the relation. This means that relative temporal expressions such

feature	values	comments
rel		nature of referent
	doc	expression is relative to document
	event	expression is relative to an event
conn		temporal connective
	>	future
	<	past
	=	now
amnt		an amount (measured in units)
	number	a number
unit		a unit of temporal measurement (quantified in amnts)
	hour	
	day	
	week	
	month	
	year	
daypart		a sub-part of a day
	morning	
	afternoon	
	evening	
	night	
hour		an hour expression
	HHHH	a numeric string in 24-hour format
day		day of the week
	mon	Monday
	⋮	⋮
	sun	Sunday
dd		day of the month
	DD	a numeric string between 1 and 31
mm		month of the year
	MM	a numeric string between 1 and 12
yy		year
	CCYY	a numeric string between 1900 and 2099
mod		a modifier
	early	
	late	
abs		absolute date value
abs:day	number	number of days since Jan 1st 1987
abs:wk	number	number of weeks since Jan 1st 1987
zone		time zone
	gmt	Greenwich Mean Time
	est	Eastern Standard Time

Table 4.6: Typology of temporal semantics in the TAM

as '3 days ago' can be represented as:

$$\begin{bmatrix} \text{rule: 4r} & \\ \text{sem:} & \begin{bmatrix} \text{amnt: 3} \\ \text{unit: day} \\ \text{conn: } < \\ \text{rel:doc} \end{bmatrix} \end{bmatrix}$$

`daypart`

Temporal expressions frequently make reference to parts of days — for example, 'friday morning' and 'last night'. Although in this latter case it appears that `night` is functioning in the same way as, say, `week` (i.e. as a `unit`), semantically `night` in this context carries both a `unit` value of `day` and a `daypart` value of `night` — meaning that it is synonymous with 'yesterday night'. This means that '3 nights ago' differs only slightly from the example above:

$$\begin{bmatrix} \text{rule: 4r} & \\ \text{sem:} & \begin{bmatrix} \text{amnt: 3} \\ \text{unit: day} \\ \text{daypart: night} \\ \text{conn: } < \\ \text{rel:doc} \end{bmatrix} \end{bmatrix}$$

`hour`

Hours are represented in four digit 24 hour format, as used in the corpus.

`day`

The day of the week is represented in abbreviated form, e.g. `sat`.

`dd`
`mm`
`yy`

`dd`, `mm` and `yy` contain the date, month and year in two digit numeric format.

`mod`

Only the two modifiers `early` and `late` are used. It is not clear whether the inclusion of these is useful.

`abs`

The absolute day `abs:day` and week `abs:wk` numbers referred to by the temporal expression, where possible.

`zone`

Time zone, if mentioned. Unlikely in practice to be a distinguishing feature between two temporal expressions.

4.6.2 Knowledge sources for temporal analysis

The primary task of the semantic analysis component within the TAM is to interpret the representations produced by parsing with respect to the knowledge base, resulting in a fully regularised, explicit date representation.

The knowledge base used in the TAM consists of an overt calendar data structure converted from the output of the Unix `cal` command into Prolog facts. Each fact has the format

```
cal(CCYY,MM,CW,D_List,CD_List).
```

where CCYY is the year, MM the month represented in numeric form (between 1 and 12), CW the cumulative week number, D_List an ordered list containing the dates in that week for days Sunday through to Saturday, and CD_List a similarly ordered list with corresponding cumulative day numbers. For example:

```
cal(1989,9,142,[10,11,12,13,14,15,16],[984,985,986,987,988,989,990]).
```

Having explicit representations of day numbers, week numbers and days of the week means that it is trivial to compute relative dates within the Prolog framework of the TAM.

4.6.3 Relative temporal phrases

The primary purpose of the knowledge base within the TAM is for the resolution of relative temporal phrases to explicit, absolute expressions. Although anaphora resolution in general is a difficult task, and one for which a variety of approaches have been proposed (see [Hobbs 86] and [Sidner 81] for previous work on anaphora resolution), the problem is greatly simplified within the domain of temporal anaphora. This is due to two factors: potential temporal referents (i.e. expressions) are less numerous than nominal referents (expressions); and temporal expressions are easy to identify.

We classify all temporal anaphora into two classes. The first includes those expressions that make reference to a previously mentioned **event** time. The second includes expressions that refer to the **document** time. The examples given below refer to a previously mentioned event time.

$$\text{three days}\left\{\begin{array}{l}\text{later}\\ \text{after}\\ \text{previously}\\ \text{before}\\ \text{earlier}\end{array}\right\}$$

The range of expressions that refer to the document date is much wider. The set of examples below illustrates some of these.

$$\left\{\begin{array}{r}\text{this}\\ \text{that}\\ \text{next}\\ \text{last}\\ \text{the following}\\ \text{the past}\end{array}\right\}\text{week}$$

three days ago

yesterday

tomorrow[5]

Relative temporal phrases are encoded using the set of temporal connectives contained in table 4.6. As an example of how such phrases are resolved, we shall consider the phrase 'early yesterday afternoon at around 1430 gmt'. This is parsed to give the semantic representation (left) and, after interpretation, the resolved form (right) shown below.

[5] As an aside, Hindi has a single word *kal* (pronounced *cull*) for referring to days either side of the current day — ambiguity is resolved through the use of tense. The duality that exists in English may contribute to the level of redundancy in the language, but it does make our approach to temporal analysis easier.

$$\begin{bmatrix} \text{rule: r1i} & \\ \text{sem:} & \begin{bmatrix} \text{mod: early} \\ \text{unit: day} \\ \text{conn: } < \\ \text{amnt: 1} \\ \text{rel: doc} \\ \text{daypart: afternoon} \\ \text{hour: 1430} \\ \text{zone: gmt} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \text{dd:19} \\ \text{mm:11} \\ \text{yy:1989} \\ \text{abs: } \begin{bmatrix} \text{wk:152} \\ \text{day:1054} \end{bmatrix} \\ \text{mod:early} \\ \text{daypart:afternoon} \\ \text{hour:1430} \\ \text{zone:gmt} \\ \text{rel:doc} \end{bmatrix}$$

Because of the event/document temporal dichotomy, the sentence processor maintains a pair of current potential referents for the semantic interpreter — hence the feedback loop in the TAM architecture shown in figure 4.1.

The forms referring to document date are traditionally thought of as being deictic rather than anaphoric — as they would be here, were it not for the fact that we have the document date (the context relied upon in deixis) explicitly available. The classification of referring expressions into event- and document-based is therefore analogous to that of anaphora and deixis. The grouping of the two classes under the heading of anaphora is not important — from a theoretical point of view, it is arguable whether the header (containing the document date) is a part of the document itself or not; and from a practical viewpoint, the two classes of references are functionally isomorphic, using the same resolution mechanism and the same underlying representations.

Finally, while this section has discussed temporal anaphora, it has not mentioned other forms of reference such as cataphora. Temporal cataphora, although present in other forms of narrative texts such as novels (and often used as a narrative technique in visual media [Tarantino 92]), are very rare in news texts.

4.7 Temporal analysis illustrated

This section will follow a sample text through the TAM, from preprocessed text to constraints. We shall use the same text as used in table 3.3 (see section 3.5). It is reproduced here in table 4.7 with keywords and corresponding temporal phrases marked up as follows:

- *`WORD`* — a keyword;
- $\boxed{\texttt{WORD}\ \textit{WORD}}$ — a parsed string;

S	C	Text
0	1	Bogota,
0	2	[[4 *Feb* 90]] (acan efe).
1	1	A bomb exploded [[*yesterday*] in] downtown Aracataca.
2	1	Police in Bogota say that the JPF were responsible for the attack.
3	1	Several buildings were damaged in the blast.
4	1	In a similar incident [[ten *days* ago]],
4	2	the JPF attacked an army installation in the town of Rivera.
5	1	The guards shot two terrorists.
6	1	JPF guerillas are known to frequent the town.
7	1	[[*Saturday*]]'s attack occurred [[at around *1725*]] local time,
7	2	and [*may*] have been timed to disrupt rush [*hour*] traffic.

Table 4.7: Example input text as viewed by the TAM

- [[WORD *WORD*] WORD] — the extent of the largest string with which the parser was initially presented.

The TAM finds seven keywords, five of which produce temporal phrases. These are shown in table 4.8, along with their sentence (**s**) and clause (**c**) numbers, the **document** and **event** time at the point the **phrase** is found, the **semantics** produced by syntactic analysis, and the **interpreted** form resulting from semantic analysis.

At the end of processing, the interpreted semantic forms are analysed in order to determine which pairs of expressions are incompatible and which constraints, therefore, should be output to the event manager. Compatibility between structures is defined using simple feature unification. The final constraints are represented as Prolog facts using the predicate `constraint/5`, which has the syntax `constraint(Document, Module, [Sentence`$_1$`, Clause`$_1$`], [Sentence`$_2$`, Clause`$_2$`], Arity`[6]`)`. In the case of our example, five constraints are found.

```
constraint(txt01,time,[0,2],[1,1],1).
constraint(txt01,time,[0,2],[4,1],1).
constraint(txt01,time,[0,2],[7,1],1).
constraint(txt01,time,[1,1],[4,1],1).
constraint(txt01,time,[4,1],[7,1],1).
```

[6] Always `1` at present, and therefore redundant. A value of `-1` would signify that there is evidence *against* a constraint between the two clauses.

s	c	phrase	document	event	semantics	interpreted
0	2	4 feb 90	—	—	$\begin{bmatrix}\text{rule: r11} & \\ \text{sem:} & \begin{bmatrix}\text{dd: 4}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 4} & \\ \text{mm: 2} & \\ \text{yy: 1990} & \\ \text{day: sun} & \\ \text{abs:} & \begin{bmatrix}\text{wk: 163}\\ \text{day: 1131}\end{bmatrix}\end{bmatrix}$
1	1	yesterday	$\begin{bmatrix}\text{dd: 4}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 4}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{rule: r16} & \\ \text{sem:} & \begin{bmatrix}\text{unit: day}\\ \text{conn: <}\end{bmatrix}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 3} & \\ \text{mm: 2} & \\ \text{yy: 1990} & \\ \text{abs:} & \begin{bmatrix}\text{wk: 162}\\ \text{day: 1130}\end{bmatrix}\\ \text{rel:doc} & \end{bmatrix}$
4	1	ten days ago	$\begin{bmatrix}\text{dd: 4}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 3}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{rule: r4} & \\ \text{sem:} & \begin{bmatrix}\text{amnt: 10}\\ \text{unit: day}\\ \text{conn: <}\\ \text{rel: doc}\end{bmatrix}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 25} & \\ \text{mm: 1} & \\ \text{yy: 1990} & \\ \text{abs:} & \begin{bmatrix}\text{wk: 161}\\ \text{day: 1121}\end{bmatrix}\\ \text{rel: doc} & \end{bmatrix}$
7	1	saturday	$\begin{bmatrix}\text{dd: 4}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 25}\\ \text{mm: 1}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{rule: r2ii} & \\ \text{sem:} & \begin{bmatrix}\text{amnt: 1}\\ \text{rel: doc}\\ \text{unit: week}\\ \text{conn: <}\\ \text{day: sat}\end{bmatrix}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 3} & \\ \text{mm: 2} & \\ \text{yy: 1990} & \\ \text{abs:} & \begin{bmatrix}\text{wk: 162}\\ \text{day: 1130}\end{bmatrix}\\ \text{rel: doc} & \end{bmatrix}$
7	1	at around 1725	$\begin{bmatrix}\text{dd: 4}\\ \text{mm: 2}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{dd: 25}\\ \text{mm: 1}\\ \text{yy: 1990}\end{bmatrix}$	$\begin{bmatrix}\text{rule: r10} & \\ \text{sem:} & \begin{bmatrix}\text{hour: 1725}\\ \text{daypart: X}\end{bmatrix}\end{bmatrix}$	$\begin{bmatrix}\text{hour: 1725}\\ \text{daypart: X}\end{bmatrix}$

Table 4.8: Interpretation of temporal phrases

These constraints specify, respectively, that the clause containing '4 feb 90' cannot refer to the same event as the clause in the first sentence containing the phrase 'yesterday', the clause in sentence four containing the phrase 'ten days ago' or the clause in the final sentence containing 'saturday'; that a constraint exists between the first clause of sentence one and the clause in sentence four; and that there is a similar constraint between the clause in sentence four and the first clause of sentence seven.

No explicit assumptions about the event structure of the document are made at this stage — the role of the TAM is simply to identify constraint relationships between clauses in the document and pass these on to the event manager.

4.8 Summary

This chapter has presented the temporal analysis module of CONTESS, and has discussed the syntactic and semantic processing issues encountered in interpreting temporal expressions. We began by discussing some of the issues surrounding the presentation of temporal information in discourse. We also showed how a traditional view of temporal information (classified into explicit, implicit and anaphoric sources) is not a useful one for the purpose of interpreting temporal expressions in news reports. Various properties of temporal phrases were introduced. In particular, we gave a definition of what we mean by temporal phrases, showed how such phrases were identified, and discussed the use of corpora in building a set of temporal phrases.

The Temporal Analysis Module (TAM) was then introduced, and the overall architecture shown. Various aspects of the syntactic processing of temporal phrases were discussed, including keyword analysis, island-driven parsing, and the grammar of temporal phrases. We presented our typology of semantic relations, and showed how relative temporal phrases may be interpreted with reference to either a document or an event time. Finally, we gave an example of temporal analysis applied to a sample text, showing the keywords, phrases, the semantics (interpreted and uninterpreted) and the resulting constraints.

The next chapter discusses the second of CONTESS's analysis modules, the Locative Analysis Module.

Chapter 5

Locative analysis

5.1 Introduction

In terms of input/output, the locative analysis module (LAM) is identical to the TAM. It takes a preprocessed text as input, and produces a set of constraints describing, from its own point of view, what it considers to be the restrictions on the discourse structure of the text.

The LAM is also similar to the TAM in terms of internal architecture. That is, it can be represented using the diagram in figure 4.1, and shares the same sentence processor and parser as the TAM. Module-specific resources include the keywords list, grammar, knowledge base and semantic interpreter, and represent 60% of the LAM[1]. Although the design of the two modules is similar, the amount of emphasis placed on (and use made of) particular components within each module differs — the knowledge base in the TAM, for example, is simply a formal representation of a calendar; in the LAM, whilst the knowledge base plays an identical role in regularising and interpreting the parsed phrase, it is both more complex and creative in the analysis process.

The structure of this chapter reflects the symmetry that exists between the design of the TAM and the LAM, and follows a similar path to chapter 4. We begin by looking at the presence of locative information in discourse, and in particular in news texts. We continue by presenting an analysis of locative phrases, and give an overview

[1] This figure does not include the knowledge base, which is both automatically constructed and very large (see section 5.6.2). Approximately 600 lines of Prolog code are shared by the TAM and LAM, with roughly 500 and 1000 lines of module-specific code respectively.

of the LAM. We then present the LAM in greater detail; in particular, we will look at aspects of the syntactic and semantic analysis components. Finally, we illustrate locative analysis by showing how phrases in a sample text are processed.

5.2 Locative information in discourse

Although there exists a wide body of research into the analysis of spatial expressions in discourse (see for example [Rigler 92] for a theoretical exposition of spatial structure in narrative texts), the analysis of explicit locative phrases in discourse has not received as much attention as temporal phrases.

Within the context of information extraction, most systems currently do some form of locative analysis. The "named entity" extraction task of MUC-6 [ARP96] includes the recognition and marking up of "name[s] of politically or geographically defined location[s] (cities, provinces, countries, international regions, bodies of water, mountains, etc.)" [Sundheim 95].

The tripartite classification that we presented in section 4.2 for temporal phrases does not map fully onto locative phrases. Explicit and anaphoric locative expressions clearly exist; implicit locative expressions do not.

- **Explicit** — phrases such as 'in Bogota', 'throughout Libertador Municipality' and 'in a residential district of Lima' are examples of explicit locative expressions. Unlike tense in the case of temporal expressions, there are no non-phrasal sources of locative information in narrative texts.

- **Anaphoric** — anaphoric locative information sources are illustrated by expressions such as 'there' and 'that town'. Phrases of the form '100km away' and 'in the suburbs' also qualify as anaphoric expressions, as we require a referent in order to resolve their interpretations.

As we have previously suggested, implicit temporal information sources are manifested in the form of iconic sequencing — the understanding that the flow of events in a narrative reflects the order in which those events actually occurred. There does not appear to be an equivalent locative information source. Although the constant forward progression of the arrow of time is something that we all subconsciously (and implicitly) take for granted, there exists no forced spatial movement, and hence no implicit sequencing.

5.3 Locative phrases

5.3.1 Definition

The following phrases are illustrative of the class of fragments that we consider as locative phrases. The first two examples were previously used in section 4.3.1; again, they are taken from the MUC-4 development corpus, and locative phrases are highlighted in italics.

- 3 members stationed *in la union department* reported that two fmln rebels were killed and five others wounded during clashes *near el refugio farm, san miguel jurisdiction* in the afternoon of 27 november.
- leftist guerrillas early tuesday attacked a residential neighborhood *in northern san salvador* where many government and military leaders have their homes.
- [the attacks] affected the u.s.-chilean cultural institute, a mormon church in the capital, and a branch of the "citicorp" financial firm *in vina del mar, 125 km northeast of santiago.*
- a car-bomb exploded in front of the prc embassy, which is *in the lima residential district of san isidro.*

It can be seen from the above that, as with temporal phrases, locative phrases can be omitted with little or no loss of grammaticality. However, also as in the case of temporal phrases, there are a small number of constructions that rarely, if ever, occur without locative phrases — for example, the relative clause in the last example.

The above examples contain some phrases that, although we have not marked them as such, might be considered to be locative phrases, such as 'in front of the prc embassy'. This is because we have to adopt a level of granularity to work above and, in the case of the current implementation of the LAM, we have opted to restrict this to the level at which the knowledge base operates — that is, with the smallest geographical objects being ports, airports and villages. Whilst this means that the LAM is therefore unable to propose different events on the basis of incidents occurring at distinct locations below this level of granularity, this is in effect rarely a problem. The reason for this is that, at such a fine level of granularity, locative descriptions tend to include a relative macro-level description, for example 'in the south of Tulcan', that allows the LAM to construct a useful locative representation.

As was the case with temporal phrases, we require a working definition for locative phrases. The functional similarity between the two classes of phrases has been recognised previously [Creary *et al.* 89], the main difference between the two being that, whereas temporal phrases semantically link discourse events to points in time, locative phrases link discourse events to points in space.

Definition 3 ***Locative phrases:*** *overt constituents that, although syntactically optional, play a crucial role in semantically linking physical discourse events to points in space.*

5.3.2 Identification

As with the TAM, the first stage in the construction of the locative analysis module involved compiling a list of locative phrases. Using the same techniques employed in the development of the TAM, we derived a list of key location-related words for identifying locative phrases. This initial members of this list were derived through intuition and, as the body of locative phrases they retrieved began to grow, so the list itself was amended.

Further identification of locative phrases was facilitated through the use of the gazetteer from which the LAM's knowledge base is derived (see section 5.6.2). The gazetteer contains approximately 4,500 Latin American locations[2] (cities of various sizes, airports, ports, provinces, islands and countries), and as such is a rich source of information on the "words" that contribute most frequently to Latin American names.

Although 86% of the 2,967 word types in the Latin American portion of the gazetteer occur there only once, Spanish language locative descriptions contain a small number of word types (32) that have a very high token count. These words are listed in descending order of frequency in table 5.1.

[2] This represents 3.5% of the original global gazetteer made available to MUC-5 participants.

```
de, san, puerto, la, el, isla, villa, del, santa, general,
los, las, rio, jose, juan, nueva, maria, estancia, pedro,
bahia, francisco, santiago, rosa, concepcion, punta, monte,
martin, carmen, paz, cruz, colonia, antonio
```

Table 5.1: High frequency Latin American locative words

These words were added to the phrase-identification list, along with members of closed classes such as modifiers (`north, south, southwest` etc) and habitation types (`village, town, district` etc). Table 5.2 contains locative phrases representative of those extracted using this method.

department	
	in the department of Ayacucho
	in Perulapan jurisdiction, Cuscatlan department
	...
southern	
	in southern Bogota
	in La Azulita (southern Putumayo)
	...
west	
	300 km to the west of Bogota
	in western Medellin
	...
de	
	in San Carlos De Bariloche
	Rio De Janeiro (Brazil)
	...

Table 5.2: Identification of locative phrases

5.4 Overview of Locative Analysis Module

As suggested earlier, the overall architecture of the LAM is identical to that of the TAM. For this reason, we shall restrict our discussion to covering only those aspects of the LAM that differ from the TAM. These details shall be covered in greater detail in the following sections.

One of the main differences between the TAM and the LAM involves the use made of the knowledge base and the semantic interpreter. Although the knowledge base in the LAM is used for the same purpose as that in the TAM, i.e. the enrichment and interpretation of parsed phrases, the LAM's knowledge base is richer and capable of adding greater content to the interpreted phrase.

A further difference is in the use made of the preprocessed text. As stated earlier in section 3.5.3, locative proper nouns are marked up and grouped together. Implications of this are discussed in sections 5.5.1, 5.5.2 and 5.7 below.

5.5 Syntactic analysis of locative phrases

As with the TAM, the LAM consists of an island-driven parser, a list of keywords, and a grammar. As the parser is identical to that used in the TAM, we will not discuss it here.

5.5.1 Keywords

The LAM keywords list contains 60 explicit keyword forms (again, including morphological variations) and one generic keyword type.

Explicit keywords

The keywords used to identify locative phrases for grammar development (see section 5.3.2) are clearly well-suited to form the basis of the LAM's keyword list used to trigger island-driven parsing. The list of explicit keywords is shown in table 5.3.

In common with those used in the TAM, the keywords can be grouped into classes — for example, geographical descriptions (Spanish, e.g. 'puerto', 'isla', 'estancia'[3], and English, e.g. 'town', 'province', 'airport') and modifiers (e.g. usually in English, e.g. 'north' and 'southwest').

Another subset of keywords consists of Spanish function words, notably 'el', 'la', 'los',

[3] In many Latin American dialects of Spanish, *estancia* literally translated means *farm*, or *small ranch*. The meaning in Modern Spanish has evolved to signify a *living room*.

de, san, puerto, la, el, isla, villa, del, santa, general, los, las, rio, jose, juan, nueva, maria, estancia, pedro, bahia, francisco, santiago, rosa, concepcion, punta, monte, martin, carmen, paz, cruz, colonia, antonio, department, port, city, airport, municipality, province, jurisdiction, hamlet, village, town, district, suburb, region, state, country, island, north, east, west, south, northeast, southeast, northwest, southwest, northeastern, southeastern, northwestern, southwestern

Table 5.3: Locative analysis module keywords

'las', 'de' and 'del'. Spanish place names are less agglutinating than English ones — so that individual words (typically function words like 'de' and 'las', but also geographical features such as 'villa' and 'nueva') are maintained as separate morphological units, and survive the processes of assimilation and agglutination that give English-speaking countries such high number of single-word locations.[4] The list of keywords therefore includes all those high frequency locative words contained in table 5.1.

The keyword list also contains many words typically considered as personal proper names.[5] This reflects the common Latin American practice of naming locations after people (especially generals). For example, there is a district in Chile called 'Libertador General Bernardo O'Higgins'. The LAM knowledge base contains 36 distinct entries for towns in Argentina named after generals — from 'General Acha' to 'General Villegas'. This makes identifying location descriptions all the harder; the only way the LAM can distinguish between 'General Saavedra' the person and 'General Saavedra' the village in Santa Cruz, Bolivia, is by context.[6]

[4] Age and invasion no doubt play a role in the mutation of place names, both in terms of orthographic and spoken forms. As foreign-language invaders assign names to the places in their newly acquired lands, or reinterpret already existing names, it is easy to see how place names can alter and fuse together.

[5] For recent work on name class ambiguity resolution, the reader is referred to [McDonald 93], [Sundheim 95] and [Wakao *et al.* 96].

[6] Gazetteers clearly have their problems. The situation becomes much worse when presented with the full gazetteer from which our own is derived — Indonesia contains villages called *In*, *Thing* and *That* [Evans].

Locations

As with the TAM, the LAM keyword component also includes a generic form — in this case, locative proper nouns that have been marked up during preprocessing. A Perl script is used to mark up as potential place names all words in a document that appear in the Latin American gazetteer. These are only potential place names — indeed, as we have just suggested, many of the terms marked up will in fact be personal proper names.

Multiple adjacent marked up terms are combined together to form potential place name strings. For example, the sentence below is marked up as follows, with locative proper nouns shown outlined by a box.

markup: guerrillas killed a peasant in the city of $\boxed{\text{flores}}$, in the northern $\boxed{\text{el}}$ $\boxed{\text{peten}}$ department

combine: guerrillas killed a peasant in the city of $\boxed{\text{flores}}$, in the northern $\boxed{\text{el peten}}$ department

Potential locative strings are therefore treated as keywords — though not (except in special circumstances as we shall see in section 5.5.2 below) as complete locative phrases in themselves. Found in the appropriate context, however, they can be a part of a locative phrase.

5.5.2 A grammar of locative phrases

The grammar in the LAM is considerably smaller than that in the TAM. It consists of 9 phrasal rules,[7] and 81 lexical entries. As with the TAM, regular expression-style forms are used within the phrasal rules. We shall illustrate the LAM grammar by showing how the following phrase is analysed.

- lima residential district of san isidro

[7] One of the phrasal rules, listed in the grammar as rule R99, can only apply in the context of document headers. It states that a locative phrase can consist of a (marked up) locative proper noun on its own. This takes into account the idiosyncratic style of the document header; if the rule were allowed to operate in the body of the document, many personal proper names and non-positioning locations would be incorrectly interpreted as positioning locations.

The top-level rule for this phrase looks like:

```
%Grammar rule R11
R11 ---> [place#A],B,opt([of]),[place#C]:-
        R11:syn:cat <=> s,
        B:syn:cat <=> habitn,
        R11:rule <=> r11,
        R11:sem:place <=> C,
        R11:sem:type <=> B:sem:type,
        B:sem:super <=> Super,
        R11:sem:super:place <=> A,
        R11:sem:super:type <=> Super.
```

In the example phrase above, 'lima' and 'san isidro' will have been marked up as potential locations — represented as `place#[lima]` and `place#[san,isidro]`. Of the two remaining elements of the right hand side of rule R11, the first (`B` above) is a habitation type, denoted by an element with the syntactic category `habitn`. The second is an optional element, and represents simply the preposition 'of'.

Of interest in this rule are the semantic features `place`, `type` and `super`. Before describing their function, however, it will be useful to look at the structure of a lexical entry with syntactic category `habitn`.

```
X ---> [residential],[district]:-
        X:syn:cat <=> habitn,
        X:sem:type <=> suburb,
        X:sem:super <=> city.
```

A 'residential district' is defined as having semantic type `suburb` and being a subpart of something with type `city`. In the next section we will present the typology of locative features that we use and, continuing with the above locative phrase, show how phrases are interpreted using the knowledge base.

5.6 Semantic analysis of locative phrases

As is the case with the TAM, the parsing component of the LAM generates a semantic representation of the phrases that it identifies in the document. These are then interpreted with respect to a knowledge base, and a more explicit, regularised semantic structure is produced. This structure ultimately forms the basis for constraint generation, the final task of the LAM.

In this section, we begin by looking at the typology of locative semantic features used by the LAM. We then present the LAM's knowledge base, and show how its interaction with the interpreter results in the semantic output produced by the parser.

5.6.1 Typology of locative descriptions

The LAM has nine features in its semantic typology, seven of which have a fixed range of values; of the remaining two, one feature has as its value a locative proper noun, and the other an unbounded numeric. Table 5.4 presents the full set of features.

5.6.2 Knowledge sources for locative analysis

As suggested earlier, the knowledge base in the LAM consists of a fairly large (approximately 4,000 entry) gazetteer containing Latin American locations (ports, airports, cities, provinces, islands and countries) and a restricted amount of structural information (such as province, country).

The knowledge base is derived from a gazetteer that was constructed by US government analysts and provided to the Consortium for Linguistic Research, who made it available for the use of MUC-5 participants. The gazetteer contains over 4 million single line entries of the form:

> Egilsstahir (CITY 4) Sudhur-Mulasysla (PROVINCE) Iceland (COUNTRY)

where '(CITY 4)' denotes the size of the city on a decreasing scale of one to ten, 'Egilsstahir' the name of the city, 'Sudhur-Mulasysla' the province in which it is situated, and 'Iceland' the country.

The gazetteer contains a huge amount of data — too much, in fact, for our purposes. We therefore extracted from it only entries that met the following restrictions.

- all countries

feature	**values**	**comments**
type		geographical entity
	airport	
	city	
	country	
	hamlet	
	island	
	port	
	province	
	river	
	state	
	suburb	
	valley	
	village	
mod		a modifier
mod:dir	n	north
	ne	northeast
	⋮	⋮
	s	south
	cen	central
	out	outskirts
mod:dist		a distance
mod:dist:amnt	number	distance value
mod:dist:unit	km	kilometres
	m	metres
mod:pp		preposition
	in	in
	throughout	throughout
place		locative proper noun
	place	a marked up proper noun
super		properties of a containing location
super:place		
super:type		

Table 5.4: Typology of locative semantics in the LAM

- all cities with size index 1 (i.e. largest)
- all geographical entities within a country member of Latin American group[8] Argentina, Bolivia, Chile, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Peru;
- for cities within the above group, city size index must be less than 5;

The resulting mini-gazetteer was then transformed into a Prolog term `place/7` as shown below.

before: Abrego (CITY 4) Norte de Santander (PROVINCE) Colombia (COUNTRY)

after: `place([abrego],[4],[],[],[],[norte,de,santander],[colombia]).`

The seven arguments represent the entity and parameter information contained in the gazetteer — i.e. city, size, airport, port, island, province and country. Prolog term matching has been used in the knowledge base; although this results in a less readable knowledge base, we believe that the efficiency gain produced by such a representation is valuable.

The knowledge base and the gazetteer from which it is derived have an implicit structure, allowing us to abstract a hierarchy of geographical entities. This is shown in figure 5.1, and can be interpreted as meaning that countries can contain ports, provinces and islands; and that provinces can in turn contain airports and cities.

It is important to stress that the hierarchy in figure 5.1 is implicit and abstracted from the knowledge base, as opposed to explicit and implemented within it. Some aspects of the hierarchy seem strange — for example, whilst provinces can contain airports, they cannot contain ports. This is most likely due to classification methodologies adopted by the creators of the original gazetteer.

Having presented the nature of the knowledge base, and the hierarchy that has been abstracted from it, we turn now to the use that is made of the knowledge base by the

[8] As defined in [NRaD 92]. This is not an exhaustive collection of Latin American countries; rather, it includes only those which the MUC organisers deem to be "of interest".

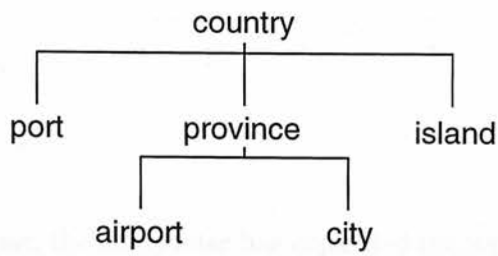


Figure 5.1: Hierarchy of geographic entities abstracted from LAM knowledge base

LAM. This can be broken down into two main areas — adding positive and negative information to the semantic representation, and ambiguity resolution. As we shall see, both functions have an effect on constraint generation.

Using the knowledge base to add content to semantic representations

Given a simple phrase such as 'in Andamarca', the parser returns an equally simple semantic representation as follows.

$$\begin{bmatrix}\text{rule: r99}\\ \text{sem: }\begin{bmatrix}\text{place: andamarca}\end{bmatrix}\end{bmatrix}$$

However, the availability of the LAM knowledge base means that further information can be inferred. The entry for Andamarca is as follows:

```
place([andamarca],[4],[],[],[],[oruro],[bolivia]).
```

This defines Andamarca as a small (size 4) city in the province of Oruro, Bolivia. Consequently, the semantic representation can be augmented to include this information.[9]

[9] A *lookup* value of *yes* denotes a successful match of the specified location in the gazetteer; a value of *no* implies that the location has not been found in the gazetteer. Although this feature-value pair plays no role in the process of constraint construction, and is thus included for informational purposes only, it is not hard to see how this feature could be incorporated into a future program that makes use of weighted constraints.

```
┌city: andamarca ┐
│size: 4         │
│airport: n/a    │
│port: n/a       │
│island: n/a     │
│province: oruro │
│country: bolivia│
└lookup: yes     ┘
```

Using the knowledge base, the interpreter has expanded the representation of the locative expression and, as a result, the LAM is in a better position to identify constraint relationships between clauses in the document. For example, if a subsequent phrase were to refer to 'the province of Santa Cruz', the LAM would be able to recognise that this is a different place to that referred to in the previous clause — on the basis that something cannot be in both Oruro province and Santa Cruz province. Were the LAM unable to use such a knowledge base, the link between the town of Andamarca and Oruro province would not have been made.

Ambiguity resolution using the knowledge base

There are actually (at least!) two Andamarcas in Latin America[10]. One of them, as we have seen above, is in Oruro province, Bolivia. The other is in Junin province, Peru. 'Andamarca' is therefore ambiguous, and the LAM must distinguish between them.

The TAM maintains a pair of temporal referents (document time and event time) for the purpose of resolving relative temporal expressions. We have already stated that the LAM interpreter is not concerned with the resolution of relative locative expressions. However, it still maintains a single locative discourse referent — initially the document location, but subsequently the event location. This is used as a context within which to interpret locative expressions.

Consequently, the above interpretation of Andamarca as being a small town in Bolivia is reliant on 'Bolivia' being available as a locative referent — either from the document header, or from a recently interpreted locative phrase. Were 'Peru' to be the current referent, the phrase would have been interpreted differently.

It is easy to see a potential problem with this approach; that is, given that the LAM

[10] There are ten cities (size 4 and up) called Santa Rosa.

is restricted to interpreting potential locative proper nouns in the context of a previous referent (always a country), how can we correctly interpret expressions denoting locations in different countries within the same document?

The answer is fairly simple; explicit references to countries and major (size 1) cities are interpreted outside of the context of the locative referent. This approach relies on the apparently widespread maxim of always introducing minor cities and other geographical entities in the context of an explicit reference to that country. This means that, while a document describing an incident in Buenos Aires (Argentina) might mention 'Bogota' (Colombia) in isolation, it would be unlikely to mention the smaller 'Caballococha' (Peru) without stating that it was in Peru.

5.7 Locative analysis illustrated

This section presents the locative analysis of a sample text. For the sake of consistency, we shall use the same example text as in the TAM illustration in section 4.7. The same system of annotation is used to represent keywords and parsed strings.

- *WORD* — a keyword;
- WORD *WORD* — a parsed string;
- [WORD *WORD* WORD] — the extent of the largest string with which the parser was initially presented.

The LAM identifies four keywords, all of which produce locative phrases. Table 5.6 shows the locative phrases, the sentence (**s**) and clause (**c**) in which they were found, and the **document** location referent. As there is no concept of an event referent in the LAM, this is not shown. The **semantics** produced by the syntactic analysis component is displayed along with the **interpreted** structure returned after the knowledge base has been accessed.

Note that the term 'Bogota' in the first paragraph of the text body is not marked up as a keyword. The reason for this lies in the use of the reported speech analysis performed during the preprocessing stage. During preprocessing, the fragment 'Police in Bogota say that' is marked up as the reporting section of a reporting act; hence, the

S	C	Text
0	1	[[*Bogota*]],
0	2	4 Feb 90 (acan efe).
1	1	A bomb exploded yesterday [[in downtown *Aracataca*]].
2	1	Police in Bogota say that the JPF were responsible for the attack.
3	1	Several buildings were damaged in the blast.
4	1	In a similar incident ten days ago,
4	2	the JPF attacked an army installation [[in the *town* of Rivera]].
5	1	The guards shot two terrorists.
6	1	JPF guerillas are known to frequent [[the *town*]].
7	1	Saturday's attack occurred at around 1725 local time,
7	2	and may have been timed to disrupt rush hour traffic.

Table 5.5: Example input text as viewed by the LAM

entire fragment is viewed by the analysis modules as a single unit (as with marked up location proper nouns), and the term 'Bogota' is therefore unavailable as a (key) word in this instance.

As with the TAM, interpreted semantic forms are analysed using feature unification to determine which constraints should be passed to the event manager. In the case of our sample text, the following four constraints are output.

```
constraint(txt01,loc,[0,1],[1,1],1).
constraint(txt01,loc,[0,1],[4,2],1).
constraint(txt01,loc,[0,1],[6,1],1).
constraint(txt01,loc,[1,1],[4,2],1).
```

These constraints stipulate that the header clause containing 'Bogota' cannot refer to the same event as the other three clauses identified as possessing locative phrases — i.e. the phrase 'in downtown Aracataca' in the first sentence; the phrase 'in the town of Rivera' in the fourth; and the phrase 'the town' in the sixth (because 'Bogota' is too big to be referred to simply as a town). Furthermore, 'downtown Aracataca' and 'Rivera' are deemed to be incompatible locations, resulting in the fourth constraint.

Again, there are no explicit event structuring assumptions made at this point. The only output from the LAM in this example is the above four constraints.

s	c	phrase	doc/event	semantics	interpreted
0	1	bogota	—	[rule: r99 sem: [place: bogota]]	[city: bogota size: 1 airport: n/a port: n/a island: n/a province: distrito especial country: colombia lookup:yes]
1	1	in downtown aracataca	[place: bogota]	[rule: r4 sem: [place: aracataca mod: [dir: cen]]]	[city: aracataca mod: [dir: cen] size: 4 airport: n/a port: n/a island: n/a province: magdalena country: colombia lookup: yes]
4	2	in the town of rivera	[place: bogota]	[rule: r1 sem: [place: rivera type: city super: province]]	[city: rivera size: 4 airport: n/a port: n/a island: n/a province: huila country: colombia lookup: yes]
6	1	the town	[place: bogota]	[rule: r12 sem: [type: city super: province place: X]]	[size: 4 airport: n/a port: n/a island: n/a country: colombia lookup: yes]

Table 5.6: Interpretation of locative phrases

5.8 Summary

In this section we have presented in detail CONTESS' Locative Analysis Module (LAM), and have discussed some of the issues involved in both the syntactic and semantic analysis of locative phrases.

We began by arguing that the tripartite (explicit, implicit, anaphoric) distinction employed in describing temporal phrases is not useful in the context of locative phrases, and that a bipartite classification (explicit, anaphoric) is more relevant.

We also looked at the nature of locative phrases in more detail — in particular how we define them, and what their function seems to be in narrative discourse. We then showed how we identified these phrases within our corpus, and discussed the usefulness of high frequency "locative words" in Spanish language place names.

The mechanisms of the LAM were introduced, in particular the syntactic and semantic analysis components. To a large extent the architecture of the LAM mirrors that of the TAM described in the previous chapter; consequently, only the elements that directly relate to the LAM were discussed — the keywords (Spanish and English), the use of marked up locative proper nouns, the grammar of locative phrases and, in particular, the function of the knowledge base in resolving ambiguity and producing regularised, explicit semantic representations.

Finally, we reintroduced the sample text first seen in section 3.5 and showed how the TAM derives constraints via semantic representations.

The next chapter presents the third and final analysis module in CONTESS, the Cue Phrase Analysis Module.

Chapter 6

Cue phrases

6.1 Introduction

It has long been recognised that cue phrases play an important role in signalling discourse boundaries [Ballard *et al.* 71]. Their explicit nature[1] makes them an attractive source of discourse structure; consequently, it is not surprising that they have been used in the past for this purpose.

The INTERPRETEXT information extraction system built by the Intelligent Text Processing group [Dahlgren *et al.* 91] contains a "discourse segmentation module" that proposes "discourse focus shifts" based on the presence of explicit segmenting connectives such as 'meanwhile' and 'in sum'.

The discourse processing module (DPM) of General Electric's NLTOOLSET (discussed previously in section 2.4.2) also makes use of cue phrases in proposing a text segmentation.

As in the case of temporal and locative information sources, however, it is very difficult to quantify (and sometimes even qualify) how useful cue phrases are for determining discourse structure. We have therefore included a Cue Phrase Analysis Module (CPAM) in CONTESS. This short chapter presents the CPAM, the third and final analysis module in our system.

[1] the quote at the front of this thesis being perhaps the most explicit discourse boundary marker of them all

6.2 Cue phrases

Cue phrases (also referred to as discourse markers [Schiffrin 87], discourse particles [Schourup 85], discourse connectives, discourse cues [Moser & Moore 95], meta-technical utterances [Zuckerman & Pearl 86], and clue words [Cohen 84]) perform a variety of functions related to the explicit indication of discourse structure. For example, *hence* and *consequently* denote inference, and *however* contrast. For the purpose of the task of flat-structure text segmentation, however, we are only interested in those cue phrases that signal a change of topic — such as new events that are in contrast or parallel to previously mentioned events.

Consequently, while English contains a large number of cue phrases (Knott's taxonomy [Knott 95] includes over 200 cue phrases), we restrict ourselves to just eighteen. Following Cohen, we can classify our confined set of cue phrases in terms of Quirk *et al.*'s taxonomy [Quirk *et al.* 72] as shown in table 6.1.[2] This list was partially constructed through inspection of the corpus, with additional phrases added from Quirk *et al.*

listing (enumerative) firstly secondly thirdly	**listing (additive, reinforcing)** what is more in addition
listing (additive, equative) correspondingly equally likewise similarly in the same way by the same token	**transitional (temporal)** meantime meanwhile
	miscellaneous in another X a further X a similar X a recent X a related X

Table 6.1: Cue phrases used in the CPAM

Clearly the cue phrases in table 6.1 do not *always* signal a potential discourse segmentation. This is due to the dual role that many cue phrases typically play — either a

[2] The subset of phrases termed **miscellaneous** are not in Quirk *et al.*'s taxonomy. (The variable X is used here only for purposes of clarity). These are domain- (or perhaps genre-) specific phrases that signal, even more explicitly than other cue phrases, that a new event is being introduced in the discourse. The quotation used in the frontispiece of this thesis can be seen as the ultimate extension of such a genre-specific discourse signal.

sentential role (for example, *similarly* used as an adverb) or a discourse role. We are only interested in cue phrases that play a discourse role within a text. Consequently, we have a problem of ambiguity resolution to address.

The issue of cue phrase ambiguity resolution in narrative texts is not an area that has received much attention. Sources of ambiguity resolution in spoken text appear to be more forthcoming — recent research [Hirschberg & Litman 93] has shown that pitch accent and prosodic phrasing are useful sources, for example.

Surface position also seems to offer some insight into the specific role played by cue phrases. Previous work in both recognition [Iwańska 91] and generation [Reichman 85] [Zuckerman & Pearl 86] [Paice 90] has assumed that early (typically initial) position in the sentence indicates a discourse use of cue phrases. This is also reflected in Schiffrin's theoretical framework [Schiffrin 87].

Although cases can be constructed containing discourse uses of cue phrases that are non-initial, surface position does seem to be a useful heuristic for disambiguation. In the next section we shall show how this heuristic is integrated into the CPAM.

6.3 Overview of Cue Phrase Analysis Module

The CPAM takes a very simple (but, as we shall see in chapter 9, quite successful) approach to cue phrase analysis. A short Perl program with access to a list of the above cue phrases searches the preprocessed document. Where such a phrase is found towards the front of (no later than four words after the start of) a sentence, a constraint is output between the current sentence and all preceding ones. That is, the presence of a cue phrase is deemed to signal a completely new event.

6.4 Cue phrase analysis illustrated

Table 6.2 shows the same example text as used in the chapters on temporal and locative analysis. As the CPAM operates by straightforward pattern matching, only cue phrases as identified are shown contained in a box.

As shown in table 6.2, only one cue phrase is identified in the text. (This is not

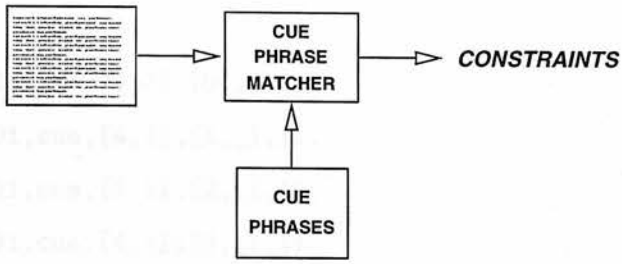


Figure 6.1: Cue Phrase Analysis Module (CPAM) architecture

S	C	Text
0	1	Bogota,
0	2	4 Feb 90 (acan efe).
1	1	A bomb exploded yesterday in downtown Aracataca.
2	1	Police in Bogota say that the JPF were responsible for the attack.
3	1	Several buildings were damaged in the blast.
4	1	In a similar incident ten days ago,
4	2	the JPF attacked an army installation in the town of Rivera.
5	1	The guards shot two terrorists.
6	1	JPF guerillas are known to frequent the town.
7	1	Saturday's attack occurred at around 1725 local time,
7	2	and may have been timed to disrupt rush hour traffic.

Table 6.2: Example input text as viewed by the CPAM

surprising; in the development corpus, only one document in seven contains a cue phrase). As it is in the fifth sentence (the fourth after the header, which is seen as sentence 0, hence the [4,1]), four constraints are output — one for each of the preceding sentences.

```
constraint(txt01,cue,[4,1],[0,_],1).
constraint(txt01,cue,[4,1],[1,_],1).
constraint(txt01,cue,[4,1],[2,_],1).
constraint(txt01,cue,[4,1],[3,_],1).
```

This states that the sentence containing the outlined cue phrase cannot refer to the same event as any of the earlier sentences. Unlike the previous two analysis modules, the TAM and the LAM, no intermediate semantic representation is produced. The constraints are simply produced on the basis of surface forms (and positions).

6.5 Summary

This chapter has presented the Cue Phrase Analysis Module, the third and most simple analysis module included in CONTESS. We began by stating that cue phrases have been used previously in information extraction systems, although no evaluation of their performance has been reported by their developers.

We then showed that the subset of cue phrases with which we are concerned is functionally constrained along two dimensions; firstly, to include only those that have a discourse (as opposed to sentential) role, and secondly, restricted to cue phrases that, in this corpus, typically signal the introduction of a new event. In conveying these restrictions, we made reference to previous uses of surface position as an indication of sentential/discourse role, and to the relation between the subset of phrases we have adopted and the taxonomy that Quirk *et al.* have used.

The simple structure of the CPAM was also presented, and its function was illustrated through the use of a example text. We also saw how the CPAM produces multiple constraints for each cue phrase it encounters (although the number of actual cue phrases is low).

This chapter concludes our presentation of the text analysis component of CONTESS. The next chapter describes the Event Manager, and shows how the constraints produced by the three analysis modules are converted into clause-event grids.

Chapter 7

Event manager

7.1 Introduction

The previous three chapters have presented the analysis components of CONTESS, self-contained modules that produce sets of constraints governing the discourse structure of input documents. However, the resulting constraints do not constitute an unambiguous structural description, and it is for this reason that we need an Event Manager (EM) to intelligently combine the constraints with various heuristics to produce a single discourse structure.

The EM can be viewed as having two distinct roles, both of which represent tasks of *identification*. The first task consists of identifying the *units* in the discourse. Although the analysis modules operate at the level of clauses, it is not necessarily the case (and indeed highly unlikely) that the discourse structure itself is best described at that level of granularity. In other words, although the constraints are identified at the level of clauses, the structure of the document is more likely to contain relationships that exist at the level of sentences or even paragraphs. One function of the EM is therefore to identify (and render perspicuous for purposes of evaluation) the level of granularity at which discourse structure is most usefully described.

As we have mentioned above, and shall discuss in more detail below, constraints alone rarely produce a single unambiguous grid description. The second task of the EM is therefore the identification of *relationships* between these units.

This chapter describes the methods used to achieve these dual identification tasks, and shows how they may be applied, in conjunction with the set of constraints $\mathcal{C}$ produced by the analysis modules, to result in grid structures. Furthermore, the modular approach taken in the construction of the analysis components will be reflected in the *controlling role* played by the EM — it is from within the EM that the constraints proposed during analysis are either selected or ignored. This is extremely important, whether it be for reasons of evaluation (analysis components and, as we shall see in section 7.3, various heuristics can be individually and jointly evaluated) or expansion (further analysis modules can be easily integrated into the system).

7.2 Ambiguity of constrained discourse structures

Figure 7.1 represents the tree of all possible grids up to a depth of 4 clauses ($n = 4$). As in figure 3.5 (on which this figure is based), shaded elements within the clause-event grids highlight elements that are shared amongst children in the tree, and are present solely as a visualisation aid.

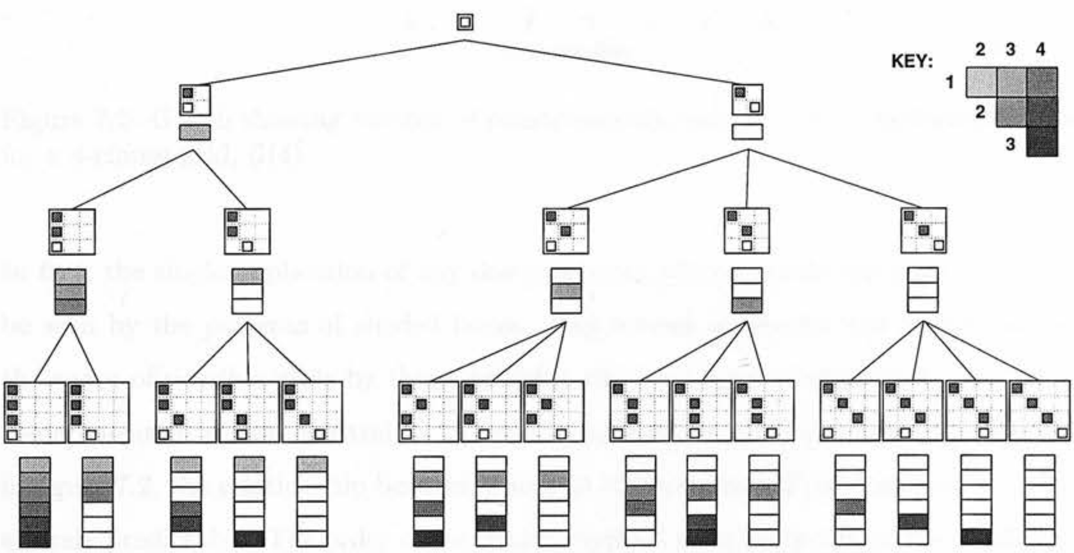


Figure 7.1: Reduction of grid possibilities through constraint application

The aim of this section is to show how the range of possible grids is affected when constraints are imposed on the relationships between elements on the grid. Superimposed on the tree in figure 7.1 is a representation of grids that are *eliminated* from the set of possible grids at each level. For example, the grid in the bottom left of the tree is ruled out in the presence of *any* constraints, i.e. when $|\mathcal{C}| > 0$. Its only sister grid is ruled out in the presence of any of the three constraints pairs[1] $\{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,3\rangle\}$ (the figure key shows how the shaded boxes denote constraints ranging from, in degrees of darkness, $\langle 1,2\rangle$ to $\langle 3,4\rangle$). Furthermore, it can be seen that the single constraint $\langle 1,2\rangle$ eliminates the five leftmost 4-clause grids (clearly, any grid in the left hand major branch of the tree will be eliminated by this constraint).

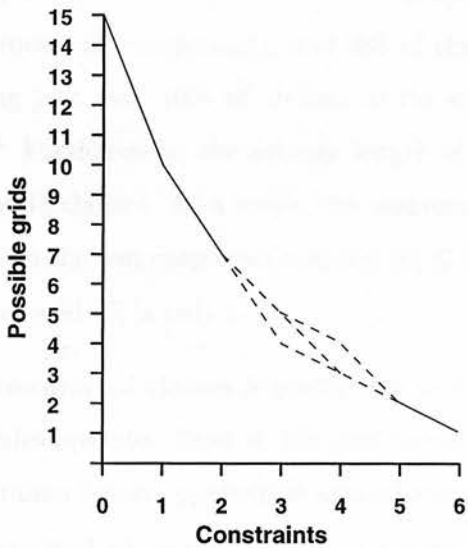


Figure 7.2: Graph showing number of constraints applied, $|\mathcal{C}|$, versus grid possibilities for a 4-clause grid, $\mathcal{G}(4)$

In fact, the single application of any one constraint will eliminate five grids — as can be seen by the patterns of shaded boxes. Any second constraint will further reduce the space of possible grids by three, bringing the total number of possible grids after the application of two constraints to seven. However, as can be seen from the graph in figure 7.2, the relationship between numbers of constraints $|\mathcal{C}|$ and grids $\mathcal{G}(4)$ is not entirely predictable. The order of constraints applied marginally affects the number of grid possibilities at each stage, resulting in three possible sequences for $\mathcal{G}(4)$. Table 7.1 gives one example set of constraints $\mathcal{C}$ for each sequence. In the table, constraints are applied incrementally from left to right.

In the 4-clause grid used in the examples above, $0 \leq |\mathcal{C}| \leq 6$. Whilst this may be useful for the purposes of illustration, in reality the proportion of constraints to clauses is

[1] The constraint notation was introduced in section 3.6.2.

$\lvert\mathcal{C}\rvert$	0	1	2	3	4	5	6	$\mathcal{C}$
$\mathcal{G}_1(4)$	15	10	7	5	3	2	1	$\{\langle 1,2\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,3\rangle, \langle 2,4\rangle, \langle 3,4\rangle\}$
$\mathcal{G}_2(4)$	15	10	7	4	3	2	1	$\{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,3\rangle, \langle 1,4\rangle, \langle 2,4\rangle, \langle 3,4\rangle\}$
$\mathcal{G}_3(4)$	15	10	7	5	4	2	1	$\{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,4\rangle, \langle 3,4\rangle, \langle 2,3\rangle, \langle 1,4\rangle\}$

Table 7.1: Table showing the effect of constraint application on grid possibilities

much lower. For example, the test corpus of 100 documents contains 3517 clauses (as defined by the preprocessing component), and 385 of these (termed *constrained*[2] clauses, and representing just over 10% of clauses in the corpus) are involved in a constraint relationship.[3] Furthermore, the average length of a text (and therefore of its corresponding grid) is 38 clauses. As a result, the maximum number of constraints per grid is also higher — in the test corpus we find $0 \leq |\mathcal{C}| \leq 52$. However, the *average* number of constraints per grid $|\bar{\mathcal{C}}|$ is only 5.

The high number of *unconstrained* clauses is partly due to the level of granularity at which the analysis modules operate. That is, the grid formed by taking into account only the identified constraints ignores potentially crucial structural aspects of the document — that clauses are part of sentences, and sentences part of paragraphs. We have therefore explored heuristics that take this into account, and present a number of them in this chapter.

7.3 Heuristics for document structuring

The sample text that we have been using in illustrating the analysis modules of CONTESS has resulted in thirteen constraints being output — five by the TAM, and four each by the CPAM and the LAM. We can combine these to show how the various aspects of the Event Manager function. Collated here are the full set of constraints produced by the analysis modules, labelled `a` to `m` for ease of reference. The original example text is repeated here in table 7.2, with columns containing sentence (**S**) and clause (**C**) numbers and corresponding fragments of text (**Text**). Sentences are shown

[2] In the case of clauses involved in constraints proposed by the CPAM, only the clause actually containing the cue phrase is deemed to be constrained.

[3] The total number of constraints found is, of course, higher. Of those clauses in constraint relationships, on average each is involved in approximately 1.381 constraints.

S	C	Text
0	1	Bogota,
0	2	4 Feb 90 (acan efe).
1	1	A bomb exploded yesterday in downtown Aracataca.
2	1	Police in Bogota say that the JPF were responsible for the attack.
3	1	Several buildings were damaged in the blast.
4	1	In a similar incident ten days ago,
4	2	the JPF attacked an army installation in the town of Rivera.
5	1	The guards shot two terrorists.
6	1	JPF guerillas are known to frequent the town.
7	1	Saturday's attack occurred at around 1725 local time,
7	2	and may have been timed to disrupt rush hour traffic.

Table 7.2: Example input text

separated by single lines; paragraph boundaries are represented as double lines.

```
a. constraint(txt01,time,[0,2],[1,1],1).
b. constraint(txt01,time,[0,2],[4,1],1).
c. constraint(txt01,time,[0,2],[7,1],1).
d. constraint(txt01,time,[1,1],[4,1],1).
e. constraint(txt01,time,[4,1],[7,1],1).
f. constraint(txt01,loc,[0,1],[1,1],1).
g. constraint(txt01,loc,[0,1],[4,2],1).
h. constraint(txt01,loc,[0,1],[6,1],1).
i. constraint(txt01,loc,[1,1],[4,2],1).
j. constraint(txt01,cue,[4,1],[0,_],1).
k. constraint(txt01,cue,[4,1],[1,_],1).
l. constraint(txt01,cue,[4,1],[2,_],1).
m. constraint(txt01,cue,[4,1],[3,_],1).
```

The method for constructing clause-event grids that we presented previously in chapter 3 relied on simply assigning each clause a position on the grid in a top-down, left-right manner, extending to new events whenever required to by the constraints. For our example text in table 7.2, this results in the grid in figure 7.3.

Here we can see that the first event shift (at S_1C_1) is due to constraint `f`. The second one, S_4C_1, is due to a combination of the constraints {`a`, `b`, `d`, `e`, `j`, `k`, `l`, `m`}, the first four being TAM constraints and the last four being CPAM constraints. Note that this particular set of constraints prevents this clause from being assigned to the second event column. A third event shift (in so far as the assignment cannot be left on the baseline) occurs in the next clause, S_4C_2. This is due to the combination of LAM

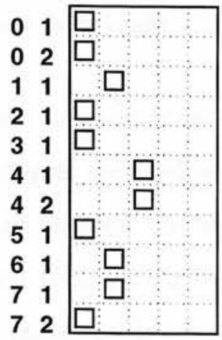


Figure 7.3: Example input text represented as a grid based purely on constraints

constraints {g, i}. Finally, clause S_6C_1 is forced from the baseline by TAM and LAM constraints {c, h}.

Even at a very simple level, there seem to be several things wrong with the grid in figure 7.3. The first three sentences, all in the first paragraph of the text body, should be aligned together on the grid. Instead, the first sentence, S_1C_1, is on a different alignment from the other two. Furthermore, the sentence in S_5C_1 would seem to describe the same event as that described by the previous sentence; again, it is misaligned. Finally, although S_7C_1 seems to have been correctly aligned with S_1C_1, the second clause in the sentence, S_7C_2 has not. (It should be pointed out that the correct clustering of S_4C_1 and S_4C_2 is due to both being individually in constraint relationships with S_1C_1 and at least one clause on the baseline.)

We can see, therefore, that although the positioning of *constrained* elements on the grid appears to be satisfactory, the positioning of *unconstrained* elements is not. Part of the solution is to perform the assignment process in two stages; initially one of constrained clauses, and then one of unconstrained clauses. We shall now describe these heuristics, beginning with those operating on unconstrained clauses (and assuming, therefore, that the constrained clauses have already been assigned to the grid). We shall then present two heuristics that operate on constrained clauses. All heuristics will be illustrated through the use of our example text.

7.3.1 Term vector frequency analysis

As the bulk of clauses in texts are unconstrained, there would seem to be a great deal of potential in exploring techniques for intelligently gelling such clauses. The first heuristic that we present (although the last implemented) involves identifying similarities between sentences[4] based on term vector frequency analysis, and constructing a "preference" list of assignment positions (such as that maintained for constrained clauses by the clustering strategies described below) on the basis of these similarities — the motivation being that, in the absence of other information (such as constraints), similarities in terms of word stems or character strings (both were investigated) might offer some insight as to the correct structure of the text.

We have implemented a version of the Smart approach to text analysis [Salton 71]. This is based on representing fragments of text as a set of weighted terms (a *term vector*). Similarities between text fragments are then calculated by analysing correspondences between term vectors. The process as applied to an individual document can be summarised as follows:

1. Remove stop words.
2. Reduce remaining words to stems.
3. Assign weights to terms.
4. Compute similarity between fragments.
5. Propose link if similarity excedes threshold.

In the first stage, words that have a high frequency in the corpus as a whole are eliminated. A stoplist of words was constructed from four main sources: a list of the most common words in the development corpus; a second list of the 100 most common words in 300 million words of Usenet traffic; a third list derived from Van Rijsbergen's book [Van Rijsbergen 79] available from the Glasgow University information retrieval

[4] Similarities are identified at the level of sentences, which then take effect on sentence initial clauses.

group[5]; and the author's intuition. The stoplist contains 339 words. After removing stop words from our example text, we are left with the following words:

```
bomb exploded yesterday downtown aracataca
police bogota say jpf responsible attack
buildings damaged blast
similar incident days ago jpf attacked army installation town rivera
guards shot terrorists
jpf guerillas known frequent town
saturday attack occurred 1725 local timed disrupt rush hour traffic
```

Remaining words are stripped of suffixes using the Porter stemming algorithm [Porter 80]. This transforms our example text into the following:

```
bomb explod yesterday downtown aracataca
polic bogota say jpf respons attack
build damag blast
similar incident day ago jpf attack army installat town rivera
guard shot terrorist
jpf guerilla known frequ town
saturday attack occur  local time disrupt rush hour traffic
```

Term weights w_{ik} are then assigned to each term T_k occurring in sentence S_i. Each sentence[6] can then be represented as a term vector of the form $S_i = (w_{i1} \ldots w_{it})$, where t represents the total number of terms in the document. We have used $tf \times idf$ (term frequency times inverse document frequency) weights, which rely on distinctive elements in sentences (i.e. those which are prominent in a particular sentence but not others) to pick out similarities between sentences. This is defined in 7.1.

$$w_{ik} = \frac{tf_{ik} \cdot \log(N/n_k)}{\sqrt{\sum_{k=1}^{t}(tf_{ik})^2 \cdot (\log(N/n_k))^2}} \tag{7.1}$$

In the above, tf_{ik} represents the frequency of term T_k in sentence S_i, N the number of sentences in the document, and n_k the number of sentences containing term T_k. The presence of the denominator ensures normalisation — otherwise long sentences,

[5] http://www.dcs.gla.ac.uk/ir/

[6] This process is typically applied to identifying similarities between *documents* in a collection, rather than *sentences* in a document. Consequences of this are discussed in section 9.9.2.

with more terms and higher frequencies, would adversely influence the detection of similarities.

Finally, the similarity between two sentences S_i and S_j is calculated using the vector similarity function in 7.2.

$$\mathrm{sim}(S_i, S_j) = \sum_{k=1}^{t} w_{ik} \cdot w_{jk} \tag{7.2}$$

For each sentence in the text, the output of the frequency analysis component consists of an ordered list of other sentences in the text that exceed a given threshold with respect to similarity ratings. Event assignment is then attempted for each (sentence-initial) unconstrained clause in turn. For our example text, the similarity judgements are as follows:

```
freq_list(txt01,1,[]).
freq_list(txt01,2,[1]).
freq_list(txt01,3,[]).
freq_list(txt01,4,[1,2,3]).
freq_list(txt01,5,[3,1,4]).
freq_list(txt01,6,[4,1,2]).
freq_list(txt01,7,[4,2,1]).
```

As can be seen above, only similarities that are scored above a preset threshold are reported. It's also clear that the judgements in this case do not reflect the actual event structure of the example text — only that of sentence 6 would result in a correct clustering, as sentences 6 and 4 *could* indeed be describing the same event. However, there are good reasons why we should not expect accurate results for this document — as we shall see in the discussion in section 9.9.2, this technique works best on larger documents.

Determination of useful threshold values, as well as term sizes (either N-grams, e.g. character trigrams, or word stems) was an area of investigation, the results of which can also be seen in the discussion towards the end of this thesis.

7.3.2 Sentence level gelling

Sentence level gelling consists of clustering together within a paragraph unconstrained clauses (as clauses are the level at which all structuring operations ultimately take place) that are in sentence initial position only, and then only when the sentence itself is not paragraph initial. That is, it only operates on clauses S_mC_n where $n = 1$ and S_m does not start a paragraph. The heuristic is concerned with (specific) clauses in the context of paragraphs — the reasoning being that paragraph boundaries correspond, at least to a small degree, to the discourse structure of the text. Given a sentence initial unconstrained clause S_mC_1 in a paragraph non-initial sentence, this heuristic assigns the clause to the same event[7] $\mathcal{E}[S_mC_1]$ as the sentence initial clause in the previous sentence, $\mathcal{E}[S_{m-1}C_1]$. This heuristic can be stated formally as shown in figure 7.4.

$$\begin{array}{ll} \text{IF} & \mathcal{E}[S_mC_1] = 0 \text{ AND} \\ & \mathcal{E}[S_{m-1}C_1] \\ \text{THEN} & \mathcal{E}[S_mC_1] = \mathcal{E}[S_{m-1}C_1] \end{array}$$

Figure 7.4: Sentence level gelling heuristic

The effect of this heuristic is seen in figure 7.5. To make small changes in grid structures easier to see, we will leave temporary assignment "shadows" in previous positions. In this case we can see that three changes have taken place. Clauses S_2C_1, S_3C_1 and S_5C_1 have all been aligned within the context of the paragraphs in which they are situated. Although in the case of our example text this heuristic has had the desired effect, it is not always the case — as we shall see in section 9.9.3.

7.3.3 Clause level gelling

The third heuristic also operates at the level of unconstrained clauses (that is, clauses that are not involved in a constraint relationship with any other clause) at any position within a sentence. It can be summarised as stating that, in the case where there exists a previously assigned clause in the current sentence (possibly appearing later in the

[7] The event assignment of clause S_mC_n is denoted by $\mathcal{E}[S_mC_n]$, which evaluates to 0 if no event has yet been assigned to that clause.

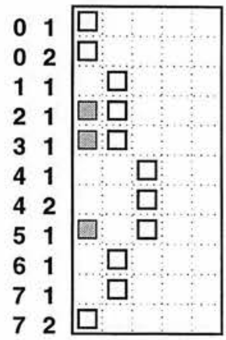


Figure 7.5: Example grid with sentence level gelling heuristics selected

sentence), the current clause S_mC_n should be given the same event assignment. Where more than one clause within the context of the current sentence already has an event assignment, the current clause should be assigned to the event of that clause which is nearest[8]. This heuristic has the effect of gelling clauses within sentences together, and is shown formally in figure 7.6.

$$
\begin{array}{ll}
\text{IF} & \mathcal{E}[S_mC_n] = 0 \text{ AND} \\
 & \mathcal{E}[S_mC_{n\pm i}] \\
\text{AND} & \\
 & \forall j \exists i (\mathcal{E}[S_mC_{n\pm j}] \rightarrow i < j) \\
\text{THEN} & \mathcal{E}[S_mC_n] = \mathcal{E}[S_mC_{n\pm i}]
\end{array}
$$

Figure 7.6: Clause level gelling heuristic

The result of running the EM on the same set of constraints with this heuristic operational is shown in figure 7.7. As we can see, only one change has taken effect — clause S_7C_2 has been assigned to the same event as the previous clause in the same sentence. As with the previous figure, shadowing has been used to highlight differences.

7.3.4 Clustering strategies

We have shown how the example grids above are constructed according to two general knowledge sources: the constraints output by the analysis modules; and the sentence and clause gelling heuristics used by the event manager. However, in building the grids,

[8] If there are two equidistant assignments available, the one earlier in the sentence is chosen.

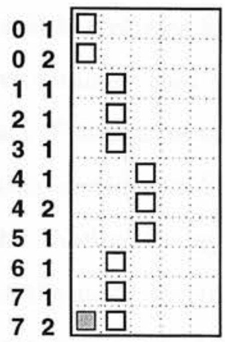


Figure 7.7: Example grid with sentence and clause gelling heuristics selected

we have made one further assumption that has so far remained hidden, and that is the strategy by which we assign *constrained* clauses that have more than one possible position on the grid.

The clause in the penultimate sentence, S_6C_1 was assigned, in figure 7.7, to the second event column. The strategy implicit in this move was to make assignments as close to the baseline as possible — and, as the constraint $\mathtt{h}$ $(\langle S_0C_1, S_6C_1\rangle)$ prevents this particular assignment from being made *on* the baseline, we have clustered it with previous sentences, S_1C_1, S_2C_1 and S_3C_1. This strategy of positioning clause-event assignments as close to the baseline as possible can be described as a *globally proximate clustering strategy*, and is presented formally for clause S_xC_y in figure 7.8. That is, it involves clustering assignments towards the baseline, which acts as a global point of reference for elements on the grid. The motivation behind this strategy is that, unless otherwise indicated (by appropriate constraints), clauses in the text are most likely to be describing the same event as that introduced at the start of the document body.[9]

LET $Ev = \mathcal{E}[S_1C_1]$
WHILE $\exists S_mC_n : \mathcal{E}[S_mC_n] = Ev$
  AND $\langle S_xC_y, S_mC_n\rangle \subset \mathcal{C}$
  DO $Ev++$
LET $\mathcal{E}[S_xC_y] = Ev$

Figure 7.8: Globally proximate clustering strategy

[9] The event introduced by the document *header* is therefore not the first choice — in fact, it is the last choice, as shall be seen in section 7.5.

This is by no means the only possible strategy for assigning ambiguous constrained clauses to the grid. An alternative strategy that we are interested in investigating is termed *locally* proximate clustering. This can be summarised as favouring assignments that result in clauses near each other on the grid being clustered together. So, instead of the global reference point of the baseline being used as a guideline for positioning S_xC_y, the local reference point of the most recently assigned (and therefore closest) element on the grid is used. The strategy is presented formally in figure 7.9.

$$
\begin{array}{l}
\text{LET } K = \langle S_0C_1 \ldots S_xC_y \rangle \\
\text{LET } i = 1 \\
\text{WHILE } \exists S_mC_n : \mathcal{E}[S_mC_n] = \mathcal{E}[K_{|K|-i}] \\
\qquad\qquad \text{AND } \langle K_{|K|-i}, S_mC_n \rangle \subset \mathcal{C} \\
\qquad\qquad \text{DO } i{+}{+} \\
\text{IF } i < |K| \text{ THEN LET } \mathcal{E}[K_{|K|}] = \mathcal{E}[K_{|K|-i}] \\
\text{ELSE LET } \mathcal{E}[K_{|K|}] = i
\end{array}
$$

Figure 7.9: Locally proximate clustering strategy

The EM keeps an ordered list of positions on the grid to which to assign clauses. The clustering strategy used therefore affects the ordering of this list and, ultimately, the shape of the grid. Using locally proximate clustering in building the grid for our example text results in the grid in figure 7.10. In this grid, the single clause sentence S_6C_1 is clustered with the immediately preceding sentence, rather than being pulled towards the baseline as with globally proximate clustering.

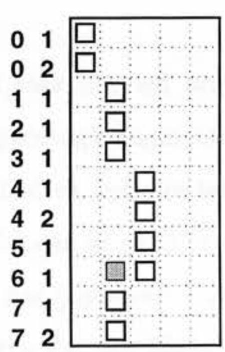


Figure 7.10: Example grid with locally proximate clustering strategy selected

7.4 Commentary, Rhetoric and Politics

The corpus that we are working with contains documents originating from a wide variety of media, including newspaper articles, TV and radio news, speech and interview transcripts, rebel communiques, and other sources [Sundheim 91]. Consequently, many documents in the corpus contain *no more than* one event.[10] Such documents can often be classed as commentary, and are typically rhetorical or political in nature. However, although these documents rarely describe multiple events, they frequently contain multiple and incompatible temporal and locative phrases, as well as instances of cue phrases.

Here is a fragment from a document that we would class as highly irrelevant:

```
TST2-MUC4-0006

   San Salvador, 22 Apr 90 (domestic service) -- [speech] [President
Alfredo Cristiani] [text] Dear radio audience: First of all, I want to
thank you for allowing me to talk to you tonight about a very
important issue for the future of our country.  Many times you have
heard us mention the word solidarity.  We have always used it as a
word that means to think about those who suffer the most, those who
have less.  Indeed, those are the people we must help, even with our
own sacrifice, and many times with a huge sacrifice by all of us.  If
love is the bond that unites family members, solidarity should be the
great pillar, the cement that unites Salvadoran society, so that,
united, we may fulfill our desire for peace, tranquility, happiness,
and well-being.
[...]
```

The inherent difficulty of the IE task has led previous researchers to develop techniques for identifying (and then ignoring) fragments of text deemed to be irrelevant.[11] These techniques have been applied at various levels — at the phrasal level in the instance of GE's NLTOOLSET [Krupka *et al.* 92]; at the sentence level as illustrated by the University of Michigan's LINK [Lytinen *et al.* 92], New York University's PROTEUS [Grishman *et al.* 92] and SRI's FASTUS [Hobbs *et al.* 92b]; at the paragraph level

[10] Using the grid representation we have described, there is conceptionally no difference between a document that contains *one* event and a document that contains *no* events. As we are fundamentally interested in representing *relationships* between events, we feel justified in adopting this position.

[11] Hobbs *et al.* claim that "in the case of the MUC-4 terrorist reports, probably only about 10% of the text is relevant."[Hobbs *et al.* 92b]

as shown by the NMSU/Brandeis system MucBruce [Cowie *et al.* 92] and SRA's Solomon [Aone *et al.* 92]; and at the document level, as illustrated by Hughes Labs' TTS text skimmer [August & Dolan 92a]. Lewis and Tong's paper in the MUC-4 proceedings [Lewis & Tong 92] presents a comprehensive analysis of text filtering in MUC systems, though focusing on the stage at which filtering is preformed (i.e. pre-, intra- or post-parsing) rather than the level at which it is applied.

In order to investigate the benefits of recognising irrelevant documents, we have implemented a relevance filter — known as the CRaP[12] filter. This component is integrated into the EM, and assigns a score to each document processed by the analysis modules. In the case of documents with CRaP scores higher than the selected threshold, the EM ignores constraints proposed by the analysis modules, resulting in minimally eventful grids for these documents.

We designed the filter to operate at the document level rather than at a lower level simply for reasons of expediency. Consequences of this decision (and a discussion of the effect of the CRaP filter in general) can be found in section 9.12.1.

Document scores are assigned based on word frequency counts for certain terms deemed to be common in irrelevant documents.[13] The presence of personal pronouns such as 'we', 'our', 'you', and 'me', modals such as 'must', 'should' and 'shall', as well as emotive terms such as 'victory' and 'urge', are taken as indications of an irrelevant document. Furthermore, rather than dividing the number of instances of these words by the number of words in the document (for purposes of normalisation), we divide by the number of distinct words in the document. This means that repetition (a common feature of rhetorical texts) also contributes towards the final score of a document.

[12] for Commentary, Rhetoric and Politics

[13] This word list was compiled by identifying irrelevant documents in the MUC-3/4 corpus (excluding the subcorpus reserved for testing of Contess) through the use of the answer templates supplied, and taking word frequency measures from the resulting body of documents.

7.5 Putting it all together — the Event Manager and parameters

There are two main reasons why the EM needs to be highly configurable. The most important of these concerns evaluation; in order to fully evaluate the performance of the analysis modules and heuristics used in discourse segmentation, we need to be able to independently specify the modules and heuristics to use.

The second reason concerns system development. In a system the size of CONTESS (over 12,000 lines of code) it is important to be able to test individual components quickly and easily. To this end, the EM contains a user interface that allows CONTESS to be configured from the command line, with various degrees of tracing available for debugging purposes. Similarly, when run in batch mode on large bodies of documents, it is easy to instruct the EM to produce grids for documents in every possible configuration. Table 7.3 lists the full set of modules and parameters available,[14] and their possible values. All items are independent of each other.

Module/Parameter	**Settings**	
Temporal analysis module	on	off
Locative analysis module	on	off
Cue phrase analysis module	on	off
Frequency analysis	on	off
Sentence gelling	on	off
Clause gelling	on	off
Clustering strategy	global	local
CRaP filter	on	off

Table 7.3: CONTESS parameters

Some examples of EM interaction are given here. For the sake of consistency, the same example text is used in all examples. Output is shown exactly as displayed on the screen; a file containing the constraints for our example text has previously been loaded into memory, and is referred to as "txt01".

[14] The CRaP filter is currently not selectable from within the EM — instead, it must be manually toggled on/off. (Other variables, such as the N-gram size and threshold used by the frequency analysis heuristic, must also be set manually.)

```
| ?- go.
manager> status.

Reading from '/hame/jeremyc/phd/manager/sys/constraints/'
OO --
 | ---,       MOD       GrID  sTLSPCFU
 | ---. <T:1 L:1 U:1> --X--> s1111111
 | ---
Perl: 1
Writing to '/hame/jeremyc/phd/eval/sys/grids/demo/com/perl/'
```

A quick word about the "graphics" in the above: "OO" means that trace mode is on (as opposed to "><"). The arrangement of lines "---" represents the status of clause and sentence heuristics; currently both are selected. The triple under "MOD" shows which of the modules are contributing constraints ("T:1" means that the TAM *is* being used, and so on; "U" is the symbol for the CPAM). The crossed arrow "--X-->" means that output is not currently being written to disk. Finally, the "s1111111" represents the configuration identifier that would be given to a grid, and denotes the selection status of each of the modules/heuristics as shown in the "sTLSPCFU" above — that is: **T**ime analysis, **L**ocation analysis, clustering **S**trategy, gelling of sentences within **P**aragraphs, **C**lause gelling, **F**requency analysis and c**U**e phrase analysis. The reason for the order of bits within this string is historical.

```
manager> display txt01.
txt01 [s1111111]
EList: [1]
EList: [1]
EList: [1]
EList: [2,1]
EList: [2,3,1]
EList: [2,3,1]
EList: [2,3,1]

0   1 2 3 4   5 6 7
1 2 1 1 1 1 2 1 1 1 2
l t t     t l   l t
+---+-----+-----+-+---
1 1
  2 2 2         2 2 2
        3 3 3
```

Having requested the EM to display the grid for text txt01, the configuration identifier (s1111111) is printed, followed by tracing information showing the element list (EList) maintained by the current clustering strategy — in this case the strategy chosen is globally proximate, referred to as strategy 1. The grid is then displayed, with horizontal lines corresponding to (from top to bottom): sentence number; clause number within current sentence; presence of a time, location or cue phrase in that clause; paragraph boundaries (+ denotes the start of a new paragraph). Below this line is the grid itself, with numbers simply acting as a visual reinforcement of event assignment — useful when examining large and complex grids.

We can now see what happens when we deselect the sentence gelling heuristic and display the grid again. The resulting grid is identical in structure to that in figure 7.7.

```
manager> sent off.
manager> status.

Reading from '/hame/jeremyc/phd/manager/sys/constraints/'
OO --
 | ---,      MOD       GrID  sTLSPCFU
 | ---. <T:1 L:1 U:1> --X--> s1110111
 |---
Perl: 1
Writing to '/hame/jeremyc/phd/eval/sys/grids/demo/com/perl/'

manager> display txt01.
txt01 [s1110111]
EList: [1]
EList: [1]
EList: [1]
EList: [2,1]
EList: [2,3,1]
EList: [2,3,1]
EList: [2,3,1]

0   1 2 3 4   5 6 7
1 2 1 1 1 1 2 1 1 1 2
l t t     t l   l t
+---+-----+-----+-+---
1 1   1 1     1
    2         2 2 2
          3 3
```

If we go back to selecting all the gelling heuristics (by turning the sentence gelling heuristic back on), and change the clustering strategy to be locally proximate, we can see that the maintained element list has changed — and so, in this example, has the grid, as we showed earlier in figure 7.10.

```
manager> sent on.
manager> strategy 2.
manager> display txt01.
txt01 [s1121111]
EList: [1]
EList: [1]
EList: [1]
EList: [2,1]
EList: [3,2,1]
EList: [3,2,1]
EList: [3,2,1]

0   1 2 3 4   5 6 7
1 2 1 1 1 1 2 1 1 1 2
l t t     t l   l t
+---+-----+-----+-+---
1 1
    2 2 2         2 2
          3 3 3 3
```

7.6 Summary

This chapter has described the Event Manager (EM), the final component of CONTESS, which accepts the constraints produced by the analysis modules and, based on the configuration selected, constructs a grid representation of the text.

We began by outlining the dual role of the EM: that of identifying the units present in the text, and the relationships that exist between them. We then showed the restriction on grid possibilities created by the application of constraints, and illustrated this phenomenon by looking at 4-clause grids. We also pointed out that, for grids of realistic sizes, the number of possible grids remains very high. The distinction was made between constrained and unconstrained clauses, and a number of heuristics were introduced that operate on each class.

The heuristic operating on constrained clauses consists of two clustering strategies — termed locally and globally proximate. Unconstrained clauses are subject to a set of gelling heuristics that function at several levels: at the level of sentences within paragraphs; clauses within sentences; and, in the case of the term vector frequency analysis component, the level of character N-grams. Each heuristic was illustrated using the example text seen in previous chapters, and grids resulting from their use were shown. Finally, we described the parametric properties of the EM, and demonstrated this through examples using the user interface.

Earlier chapters presented the analysis modules that CONTESS uses to propose constraints *restricting* the possible structure of texts. This chapter has shown how the EM, the controlling module of CONTESS, conbines these constraints with various heuristics to produce a grid *describing* the structure. Given the multitude of possible configurations that the EM can adopt, it is clearly important to be able to state which configuration is optimal. In order to do this, we must be able to evaluate its output — that is, we must be able to quantitatively compare grids. The next chapter therefore considers the task of grid evaluation.

Chapter 8

Evaluation

8.1 Introduction

In this chapter we discuss the issues involved in evaluating a system such as CONTESS. It should be clear by now, however, that we are interested in evaluating more than just a computer program — in fact, there are at least four things that we would like to evaluate.

First of all, we are interested in the possibility of evaluating grids with respect to other grids. That is, we would like to know how well a particular hypothesis grid matches a model grid. Section 8.2 presents a means of doing this, and discusses some of the strengths and weaknesses inherent in the mechanism chosen.

Of course, we are also interested in evaluating CONTESS's ability to generate good grids. Having introduced a means of measuring "goodness" in grids, we shall look at ways of quantitatively testing CONTESS. However, we would also like to know how much of a contribution is made by the individual analysis modules (and, by extension, the corresponding class of natural language fragments). In order to be sure that our discourse representation formalism is valid, and that our own segmentation is not influenced by knowledge of the program functionality, we also need to measure the agreement between naive coders, and between coders and the author, on the segmentation task. Finally, the usefulness of the various clustering parameters is also an issue to investigate.

The chapter continues by looking at previous attempts at evaluating text segmentation tools, and compares such techniques with our own approach to evaluation.

8.2 Evaluating clause-event grids

In order to be able to make objective judgements about CONTESS's performance at segmenting texts into events, we need a means of comparing clause-event grids. This in turn requires us to say what we believe is important in a grid, and what is not. For example, the particular column (i.e. event) that a clause is assigned is not important in itself — rather, it is the relation between that event assignment and those of other clauses in the document that is of interest. As an example, let us consider the two

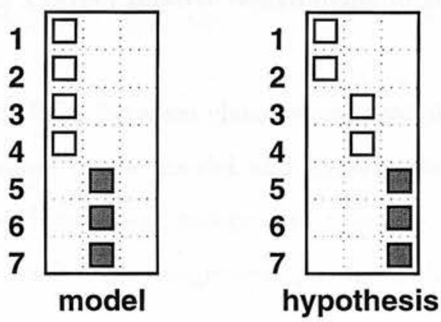


Figure 8.1: Absolute assignment position is not important

grids in figure 8.1, which describe an imaginary text. The model grid shows us that the text contains two events, the second of which is described in clauses[1] {5,6,7}. The hypothesised grid, however, proposes that there are in fact three events in the text. We can use this example to say precisely what is wrong with the hypothesised grid.

- clauses {3,4} should be clustered with clauses {1,2}

Apart from that, the hypothesised grid is an accurate representation of the model grid. The fact that the shaded boxes in clauses {5,6,7} are in the second event column in the model, and the third column in the hypothesised grid, is not relevant. The illegal grid in figure 8.2 (we use a * to denote illegal grid structures) also has clauses {5,6,7}

[1] For ease of illustration, grids in this chapter are described only in terms of *clauses*, not *sentences* and clauses as in previous chapters.

in that column. If it were not for the fact that new event assignments are always made using consecutive columns from left to right, this grid would be both legal and identical to the model grid.

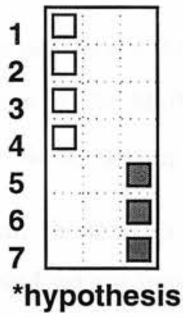
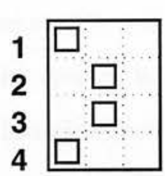


Figure 8.2: Correct relative assignments in an illegal grid

Consequently, the relationships between clause-event assignments involving these particular clauses are the same in the model and hypothesised grids. By representing grids as an ordered set of clause-event assignment relations, we can have an evaluation technique that concerns itself with assignment positions that are *relative* rather than absolute.

8.2.1 From grids to strings

To represent these relations between assignment positions, we transform grids into binary strings. These strings are built by comparing in turn the relative position of all clauses in a grid and writing a `1` if the two clauses under comparison belong to separate events (i.e. are in different columns) and a `0` if they belong to the same event. Figure 8.3 shows the resulting binary matrix for a particular four-clause grid. As the matrix exhibits symmetry across the diagonal, and assignment positions are



	1	2	3	4
1	-	1	1	0
2	1	-	0	1
3	1	0	-	1
4	0	1	1	-

Figure 8.3: Binary matrix for a four-clause grid

not compared with themselves, a grid with n clauses only requires $\frac{n^2-n}{2}$ comparisons (denoted by N) to be made, and therefore results in a binary string of that length. So, in our example grid in figure 8.3, the binary representation is the 6-character string `110011`.

This string describes the grid for an imaginary four-clause text. The average length (in number of clauses) of the documents in the MUC-3/4 development corpus is approximately 38, resulting in an average binary string length $\bar{N}$ of just over 700 characters.

Although not immediately apparent, the grid to string conversion is reversible. Given the relation between grid length and string length above, the length of the original grid can be expressed in terms of string length as a quadratic equation. The binary string can then be decoded and, observing the restrictions on grid structure given in section 3.3, the original grid is easily reconstructed (as shall be seen in figure 8.9). The relationship between grids and strings is therefore both reversible and unique.

8.2.2 Binary string comparison

Having converted our model and hypothesis grid into binary strings, the similarity $\mathcal{S}$ of the two grids can be quantified by computing the number of identically positioned ones and zeros in the two strings $\mathcal{M}$ (model) and $\mathcal{H}$ (hypothesis). This figure is then normalised by dividing by the total number of characters in the string. Multiplying the result by 100 then gives us a percentage score. This process is formally represented in 8.1, where "$\neg\oplus$" denotes a logical negated *xor*.

$$\mathcal{S}(\mathcal{M},\mathcal{H}) = \frac{\sum_{i=1}^{N}[\mathcal{M}_i]\neg\oplus[\mathcal{H}_i]}{N} * 100 \tag{8.1}$$

To illustrate this, let us consider the two grids in figure 8.4. The model grid represents a text that describes three events, whilst the hypothesised grid only contains two. The error, of course, is that the system has failed to recognise that clause {4} belongs to a new event — instead, it has clustered it with clause {1}. Consequently, one (and only one) assignment relation is incorrect — that between clause {1} and clause {4}.

This means that the binary strings will differ by one character (as can be seen in the

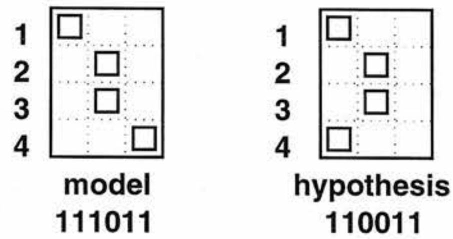


Figure 8.4: Two four-clause grids compared

figure). The string-length in this case is 6, so the similarity score of the hypothesis grid is 83.33%, as is shown in 8.2 below.

$$\mathcal{S}(\mathcal{M}, \mathcal{H}) = \frac{5}{6} * 100 = 83.33\% \tag{8.2}$$

The grids in figure 8.4 also show why it is necessary to derive binary strings of length $N = \frac{n^2-n}{2}$ rather than simply of length $N = n - 1$ (i.e. by only comparing adjacent clause-event assignments). If we were to adopt the simpler approach for the grids above, they would both result in the binary strings `101`. Restricting the examination of assignment relations to adjacent clauses clearly leads to insufficient grid descriptions.

We established in section 3.3 the relationship between grid length n and the number of possible permutations of a grid. With a small value of n (such as used above), it is possible to display and rank all grid permutations against a model grid. Figure 8.5 shows all fifteen four-clause grids ranked according to their scores against the model grid from figure 8.4.

As the grid **a** is identical to the model grid, it clearly scores 100%. Grids **b** and **c**, the second of which we also saw in figure 8.4, score 83%. They can both be said to differ from **a** in terms of one clause-event assignment relation. Grid **b** has failed to cluster clauses $\{2\}$ and $\{3\}$ together, whilst **c** has clustered $\{1\}$ and $\{4\}$ together incorrectly.

The next row contains grids that have one more assignment relation error. For example, **d** is like **b** except that it has incorrectly clustered $\{3\}$ and $\{4\}$. Grid **g** is worse than **c** for the same reason. Grid **j** also has two errors — it has clustered $\{1\}$ with $\{2\}$ and $\{3\}$. Finally, grid **o** is the worst ranking grid — it has 5 errors, as can be seen by

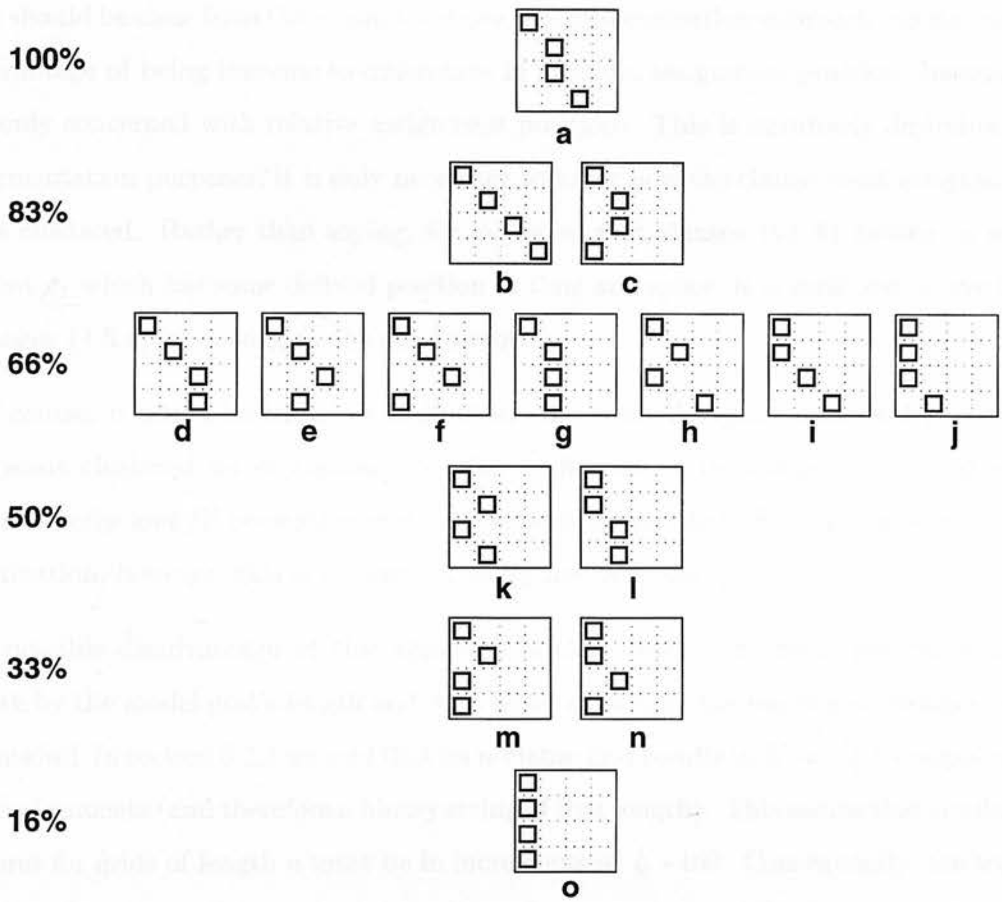


Figure 8.5: All possible four-clause grids scored against a model

comparing its binary string (`000000`, clearly) with that of the model grid (`111011`). It should be noted that the worst possible grid for a particular model will not always score 0%. The next section discusses some of the reasons for this, and suggests that this is not actually a problem in practice.

8.2.3 Benefits and drawbacks of the approach

As should be clear from the examples above, the grid evaluation approach has the major advantage of being immune to differences in absolute assignment position. Instead, it is only concerned with relative assignment positions. This is intuitively desirable; for segmentation purposes, it is only necessary to know how the clause-event assignments are clustered. Rather than saying, for example, that clauses {4,5,6} belong to some event e_1 which has some defined position in time and space, it is sufficient to say that clauses {4,5,6} all belong to the same event.

Of course, it would be trivial to find out exactly what the spatio-temporal properties of some clustered set of clauses is — the unification of the clauses' time and location descriptions (if present) would constitute precisely that. For the purpose of grid evaluation, however, this is neither necessary nor desirable.

A possible disadvantage of this approach is that evaluation scores are constrained both by the model grid's length and, to a lesser extent, by the number of events that it contains. In section 8.2.1 we said that an n-clause grid results in $N = \frac{n^2-n}{2}$ comparisons of assignments (and therefore a binary string of that length). This means that similarity scores for grids of length n must be in increments of $\frac{1}{N} * 100$. Consequently, the scores assigned to the grids in figure 8.5 are the only scores that a 4-clause grid can have.

The lowest scoring hypothesised grid will not always have a score of 0%. The reason for this is that the grid syntax usually does not allow the construction of a grid whose binary string is the exact inverse of the model. The obvious (and only) exception to this is with the maximally eventful grid (i.e. assignments along the diagonal, as in grid **b** in figure 8.5) and the minimally eventful grid (i.e. assignments straight down the vertical, as in grid **o** in the same figure). The binary strings for these are `111...` and `000...` respectively. No other inverse relationships are possible.

The graph in figure 8.6a shows the frequency distribution of the lowest scores achieved for grids of various sizes when each of the possible grids were compared against every other grid of the same dimension. (We'll call this "lowest possible score attainable for a particular grid" the grid's *baseline*). So, at one extreme each of the 5 possible three clause grids are compared against each other; at the other extreme, each of the 21,147 possible nine clause grids are compared – that's nearly half a billion comparisons for this size alone. As expected, there are only two occurrences of 0% baselines for each grid size. Note also, however, that there are occurrences of between 33% and 44% as baselines, and that the height of the highest baselines increases (although in decreasing steps) with the size of the grid (measured in clauses). This means that

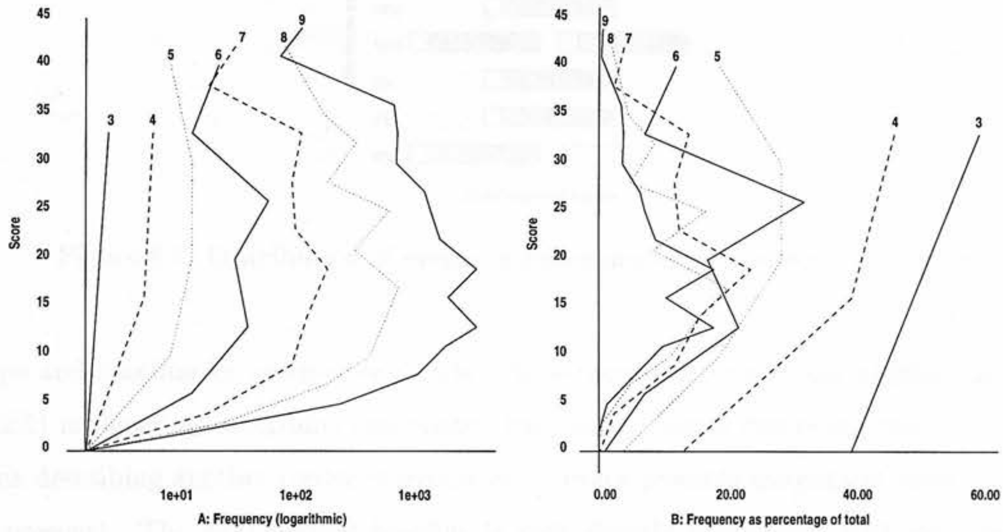


Figure 8.6: Lowest score frequency for grids of various sizes

for *certain instances* of a 7-clause grid, for example, the minimum score that it is possible to achieve when comparing that grid with all others is 42%. Clearly it is important to know how high this baseline will be for grids of realistic sizes. However, although the data represented in figure 8.6 was derived from millions of automatically created grids, the sheer combinatorics of the grids' underlying mathematics prevents us from computing the individual scores for grids longer than about 9 clauses.[2] It seems plausible that looking closely at the *kinds* of grids that result in high baselines might enable us to both predict the baselines for grids of realistic sizes, and to explain

[2] It took a Perl program running on a Sun Sparc 10 workstation a week to compare every 9-clause grid with every other 9-clause grid, a total of approximately 225 million comparisons.

why such baselines occur.

Analysis of the correlation between grid structure and baseline shows that baselines are highest when the model grid describes two events distributed across an equal number of clauses. Baselines decrease due to two reasons; primarily as the number of events being described increases, but also as the distribution of clause-event assignments becomes less even. Figure 8.7 shows, schematically, the correlation between grid structure-

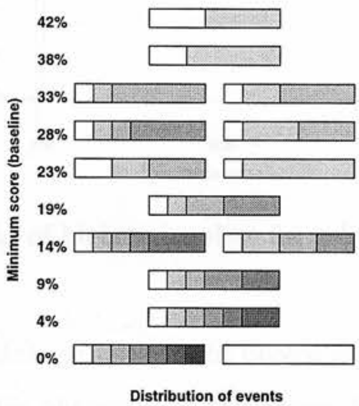


Figure 8.7: Distribution of events in minimum scoring seven-clause grids

type and baseline for seven-clause grids. The structure-type with the highest baseline (42%) is shown as describing two events, with three clauses describing one event and four describing another (order is irrelevant — every possible assignment permutation is present). The next highest baseline is split slightly less evenly, with two clauses describing one event and five describing another. On the next line, two structure-types describing three events have a baseline of 33%, and so on. The bottom line contains the maximally and minimally eventful grids (seven clauses, seven events and seven clauses, one event respectively). As explained above, they have minimum scores of 0%.

Given this information, then, we should be able to construct grids of realistic sizes that share the same *properties* as those we have observed in figure 8.7 leading to high baselines. This we did, and the results of our experiments are shown in figure 8.8.

We previously noted that, although the maximum height of baselines seemed to increase with the size of grids, the increments themselves seemed to be decreasing. On this basis we might assume that the highest point of the baselines asymptotically approaches 50% and, as figure 8.8 shows, this does indeed seem to be the case.

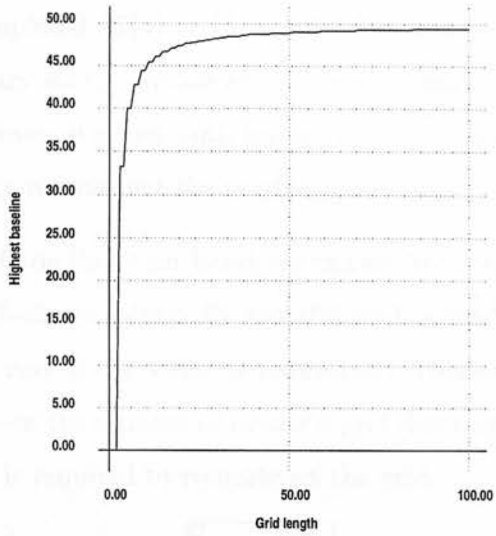


Figure 8.8: Height of highest baseline for values of n up to 100

Having identified the genus of the grids with the potential for generating high baselines, and having confirmed that this phenomenon exists even for grids of realistic sizes, it is time to consider why grids of this genus generate such baselines. Figure 8.9 represents a 5-clause, 5-event (and therefore maximally eventful) grid in several stages of definition. Attached to each grid is a side box that contains the components of the grids' binary string, with each line of the binary string having been generated by the corresponding clause on the grid. String components in emphasised font denote parts of the string that are used to construct the attached grid. In the same way, the clause-event assignments (squares) on the grid in bold are entirely derivable given the string component that is highlighted. Squares not in bold font correspond to *potential* positions of clause-event assignments.

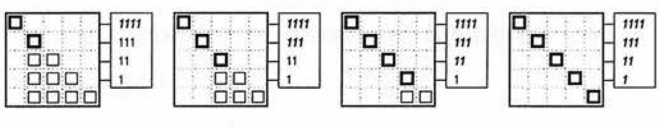


Figure 8.9: Incremental completion of maximally eventful grid, with corresponding binary strings.

For example, in the first grid in figure 8.9 (reading from left to right), only the first two lines of the grid are predictable given just the first part of the binary string. It can be seen that, given only this small part of the string, there are 15 possible ways that this grid could be completed ($\mathcal{G}(4) = 15$, see 3.4 for details). When we use another component of the binary string, we are able to predict more accurately the nature of the original grid. However, it is not until we use all the components of the string that we are able to perfectly reconstruct the original grid.

If we look at figure 8.10, on the other hand, we can see that only one line of the binary string derived from a 5-clause minimally eventful grid is needed to perfectly describe the original grid. The rest of the string is redundant. This suggests that there might be a relationship between the number of events a grid describes, and the proportion of the binary string that is required to reconstruct the grid.

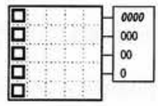


Figure 8.10: Minimally eventful grid, with corresponding binary string.

In figures 8.9 and 8.10 the number of contributions required has been equal to the number of events in the grid. In fact, the number required can be less than the number of events. Figure 8.11 illustrates this point with a 3-event grid requiring only two string components for a full, explicit description of the original grid.

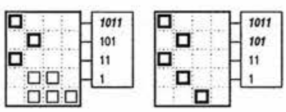


Figure 8.11: An e-event grid described using $e-1$ string components.

Further experiments appear to validate this hypothesis. The relationship between the number of events in a grid and the proportion of the binary grid needed to represent it can be described as follows:

> A grid with e events can be fully, explicitly and minimally represented as a binary string using the string contributions from no more than e clause-event assignments.

The result of this is that complete binary strings for grids with multiple, but low, numbers of events contain a greater level of redundancy than those with either single or multiple, but larger, numbers of events. Consequently, the greater the level of redundancy in a binary string, the higher the baseline for the corresponding grid will be.

In practice neither the constraint on score increments nor that on baselines is a serious one. As the average grid length is 38 clauses, the score increment is approximately 2.5%. Also, experiments involving randomly generating grids of this scale suggest that small scores are indeed quite easily achieved.[3]

Furthermore, when we represent the frequency of baselines in terms of percentages as in figure 8.6b, we can see that although the actual value of the highest baseline appears to be increasing, the proportion of grids that result in these high baselines (expressed on the horizontal axis) decreases rapidly. Consequently, the bulk of the baseline scores can be expected to remain low.

8.2.4 Alternative evaluation techniques

Recall and precision, two related metrics traditionally used in the information retrieval community, have also been used in evaluating information extraction systems [Chinchor 91]. Given that we are undertaking a kind of IE task here, recall and precision might seem obvious metrics to use in evaluating CONTESS.[4] Indeed, one very recent evaluation scheme [Vilain *et al.* 96] uses recall and precision for a similar task to that which we are interested in. Definitions of recall and precision are given below.

$$\text{recall} = \frac{\text{number correct}}{\text{possible correct}} \tag{8.3}$$

$$\text{precision} = \frac{\text{number correct}}{\text{number found}} \tag{8.4}$$

Recall measures the completeness of a system's output; precision represents the sys-

[3] It should be pointed out, however, that random grids are not the best way of testing this hypothesis. If we define a random grid as one that is constructed in such a way that, at each line (clause) in the grid, there is an equal chance of the next event assignment being made to any of the previously assigned event positions or a new one, then these random grids are not representative of the "average" grid — they are far too eventful.

[4] Ed Hovy, personal communication

tem's accuracy. The scoring scheme devised by Vilain *et al* is an interesting one, being based on the model theory of the coreference[5] links (in our framework, event pairings) in a text. The scheme works by comparing the equivalence classes defined by the links in the model and the hypothesis, and results in a computation of recall and precision that is quite elegant.

The model grid in 8.4 contains one pairing (or link), and that it between the second and third clauses. If we represent the set of pairings as <`b-c`>, then the set of pairings for the hypothesis grid is <`a-d b-c`>. Vilain *et al* define recall as being the least number of pairings that need to be added to the hypothesis in order to align the equivalence classes, and, conversely, precision as the least number of pairings that need to be added to the model in order to align the classes.

Table 8.1 shows the recall and precision values for the hypothesis grid in figure 8.4 when compared against the model grid in the same diagram. The component variables — number correct (**C**), number found (**F**) and possible correct (**P**) — are also included. Furthermore, to show that the values obtained are not reliant on the binary string construction method we have adopted, values are also given for comparisons between naive grid descriptions. The values shown, listed in order of descending naivety, are for:

- **absolute** assignment positions, represented as grid coordinates ($N = n$);
- binary string describing **adjacent** relations ($N = n - 1$);
- binary string describing asymmetric **exhaustive** relations ($N = \frac{n^2-n}{2}$);
- **model-theoretic** description following Vilain *et al.*

Description	Model $\mathcal{M}$	Hypoth $\mathcal{H}$	C	F	P	Rec	Pre
absolute	a1 b2 c2 d3	a1 b2 c2 d1	3	4	4	75%	75%
adjacent	`101`	`101`	3	3	3	100%	100%
exhaustive	`111011`	`110011`	5	6	6	83%	83%
model-theoretic	`b-c`	`a-d b-c`	1	2	1	100%	50%

Table 8.1: Grids compared using precision and recall

[5] in particular, NP coreference

We can see from the table that, in the case of the first three evaluation schemes, the values for F and P are the same. This is important because, as denominators, these are the only differentiae between recall and precision. The reason for the identity relation is simple; in evaluating grids, we always know how many clause-event assignments (or relations) we're going to find.[6] F will always be equal to N (which in turn depends on the number of clauses in the grid and the grid description we're using). The number of possible correct assignments, P, will clearly be the same. As a result, precision and recall measures for these three metrics will always give the same value.

In the case of the model-theoretic scheme, we arrive at recall and precision values that not only differ, but also appear to give intuitive results. However, closer inspection of recall and precision scores for a variety of grids shows that this scheme favours grids with spurious pairings over those with missing pairings. For example, in figure 8.5, grid `c` (recall 100%, precision 50%) would be scored higher (whatever combination of recall and precision is used) than grid `b` (recall and precision both 0%), whereas with our scheme these grids are scored equally.

The metric that we have chosen intuitively corresponds most closely to recall (the formula for similarity $\mathcal{S}$ in 8.1 is essentially identical to the recall definition in 8.3). However, the exhaustive representation of grid relations as binary strings means that we are calculating something much more meaningful than simply recall — we have a metric that focuses purely on internal relational (rather than absolute) isomorphism. This metric would seem to lend itself well to similar evaluation tasks where the relations between system decisions are both important and predictable in number.

8.3 System evaluation

Having established a means for quantitatively comparing grids, we now turn to the process of evaluating CONTESS and the linguistic information that it uses. Figure 8.12 shows an overview of the two evaluation strands that take place. The three processes that make up this evaluation, visible in the diagram as the three horizontal sections flowing from the stylised document on the left, are described in the following three

---

6 We can't just count events found — if we only count *completely* correct events, then the numerator in both formulae will be low or zero for long texts.

sections.

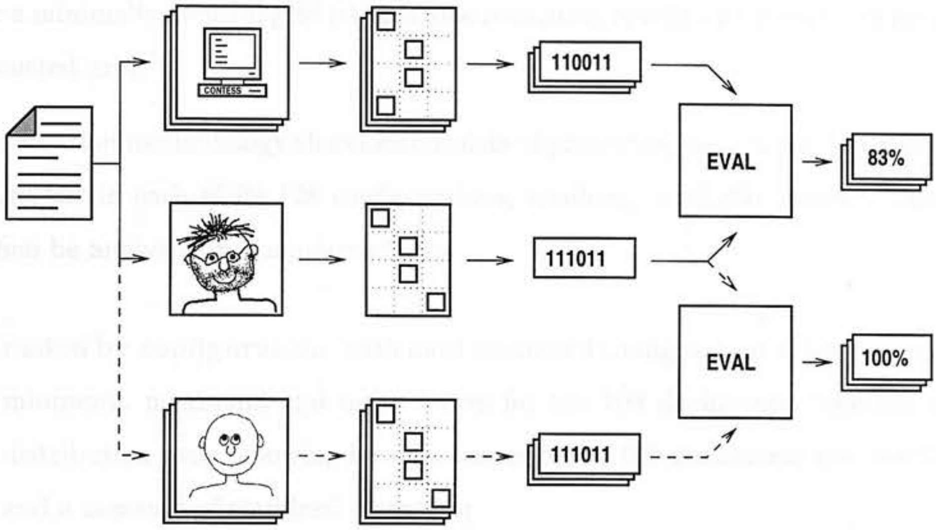


Figure 8.12: Evaluation flowchart

The documents used for evaluation come from the MUC-3/4 development and test corpus, a body of nearly 1700 texts (see section 3.2 for details). A corpus of 100 documents, disjoint from the training corpus, was selected at random from the test component of this larger body of texts. No significant system development took place after the adoption of the test corpus.

8.3.1 Methodology

The simplest way of evaluating CONTESS's performance is to have it process 100 documents and then compare the grids it generates with a set of corresponding manually-derived grids. The scores are then averaged, resulting in a figure that represents CONTESS's ability to correctly segment texts.

However, as is described in section 7, the event manager of CONTESS can be configured to selectively ignore information provided by the analysis modules and the clustering heuristics. This means that it is possible to not only evaluate CONTESS as a text segmentation tool, but also to say which kinds of information are useful in automatically segmenting texts.

As well as the scores produced by (a particular configuration of) CONTESS for a particular document, we compute the score that would be attained were a system to propose either a minimally eventful grid (that is, one consisting of only one event) or a randomly constructed grid.[7]

The evaluation methodology therefore consists of presenting each of the 100 documents to ConTESS in each of its 128 configurations, resulting in 12,800 scores.[8] This data can then be analysed in a number of ways:

- ranked **by configuration**, with most successful configuration at the top, showing minimum, maximum and mean scores for the 100 documents, together with a distribution scale (showing how the scores of the 100 documents are distributed) and a measure of standard deviation;
- listed **by document**, showing for each document the minimum and maximum scores achieved by a configuration, the scores attained by the 10 overall highest-scoring configurations, a description of the nature of the configuration that scored highest for that document if it was not one of the top 10, and finally a score produced by the minimally eventful and random grid generators;
- in terms of **component usefulness** of the analysis modules and gelling and clustering heuristics, by repeatedly grouping the configurations into two groups — those with a particular module/parameter activated, and an aligned group containing configurations with the same module/parameter deactivated — and comparing pairs of configurations. Whereas the two analyses above can be viewed as essentially black box evaluations, this third form of system analysis represents a **glass box evaluation**.

Four evaluations in total will be performed — one using the original TAM and LAM grammars and the development corpus; a second with modified grammars following rule evaluation (described below) using the same development corpus; a third with

[7] The score that would be achieved by a system proposing a maximally eventful grid (i.e. one with as many events as clauses) is simply 100 minus that scored by the system advocating minimally eventful grids.

[8] ConTESS takes approximately 90 minutes to process 100 documents in each of the 128 configurations. The scoring process then takes a further 20 minutes. Performance was measured on a Sun Sparcstation 4.

the CRaP document filter applied to the corpus; and a fourth evaluation using an "unseen"[9] test corpus. A Perl program was written that reads in and analyses a raw score file, presenting the results in each of the above ways. This is discussed in more detail in section 9.1.

8.3.2 Manual segmentation

In order to evaluate the grids that CONTESS proposes, we need a set of model grids. This was created by manually segmenting the texts, a process that is as dull as it is labour intensive. Section 8.3.3 discusses some further possible problem areas with manual segmentation.

The manual segmentation process typically involves the following steps:

- an initial skim of the whole text;
- a clause-by-clause analysis of the text, sequentially assigning clauses to events using the same grid construction rules as used by CONTESS (i.e. two clauses cannot be assigned to the same event if it is deemed that they describe different events, and event assignment is done from left to right[10]);
- a post-analysis check to verify correct clause-event assignment.

As the texts that CONTESS processes have been slightly preprocessed, the manual segmentation must take account of this. Consequently, the version of the texts that we are presented with for segmentation reflect this preprocessing. Figure 8.13 shows part of a text that has been manually segmented. The top line represents the header information, with each clause displayed on a new line along with its sentence and clause number and an associated grid that is shaded in to show event assignment. Paragraph breaks are also shown.

A user interface was designed for the input, maintenance and visualization of manually segmented texts. Initially, an HTML (HyperText Markup Language) representation

[9] The author manually segmented this corpus after all system development had been completed.

[10] As we pointed out in section 8.2, absolute assignment is immaterial — only relative assignment matters. However, we retain this order of assignment for ease of manual grid construction.

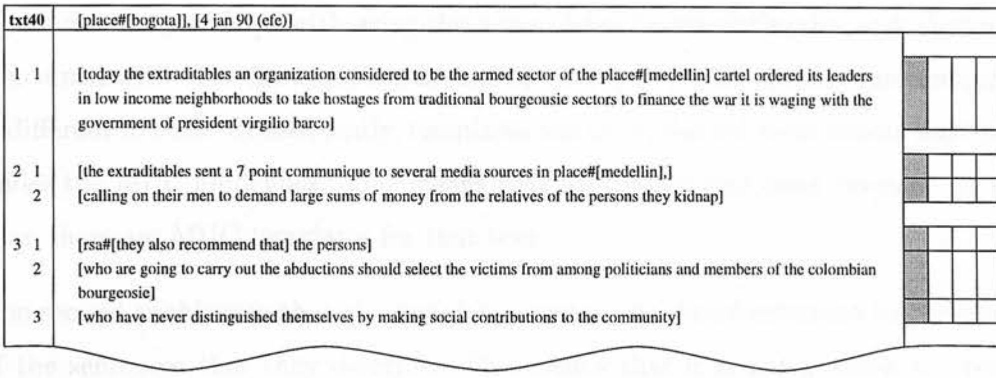


Figure 8.13: A fragment of manually segmented text

of each text in the corpus is derived from the preprocessed corpus. Clauses in the text are then displayed along side an (initially blank) "grid", and can be assigned positions on the grid by the user. Upon finishing with a particular document, the manual grids database is modified accordingly through the use of a CGI (Common Gateway Interface) program. The newly updated database then effects changes in the HTML source of the document (which is otherwise only maintained in the memory of the Web browser), completing the cycle. This process is illustrated in figure 8.14.

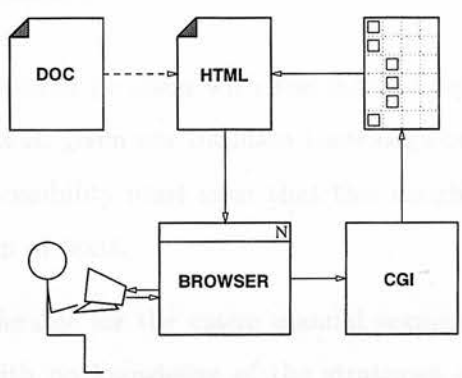


Figure 8.14: A system for maintenance and visualization of manually segmented texts

The labour intensive nature of the manual segmentation process led us to consider other, more automated, means of evaluation. One option that we looked at was the possibility of using the manually constructed templates made available as part of the MUC-3/4 development corpus. These templates constitute answer keys for the information extraction task, with one template being created for each event found in the text.

There are two problems with using these templates for the evaluation task that we are concerned with. The first problem is that the MUC notion of an event (see section 1.2) is different to ours. Consequently, templates will only exist for those events that qualify under the MUC guidelines. This means that we usually find more "events" in a text than there are MUC templates for that text.

The second problem is that the templates contain no direct reference to the location of the sentences that they describe. This means that it is not possible to "reverse-engineer" a text segmentation based on the templates. We have looked (as have Cowie *et al* [Cowie *et al.* 92], though for other reasons) at using the direct quotations given as fillers to some of the template slots as entry points into the text. However, the paucity of such direct quotations combined with frequent multiple instances in the text of those fragments that are quoted, makes this process unsuitable for deriving text segmentations. At best, it would only be possible to use the templates as a way of determining the *minimum* number of events in a text, together with the time and location of those events.

8.3.3 Human correlation

There is an obvious potential problem with the manual segmentation aspect of the evaluation process. That is, given our intimate knowledge of CONTESS's approach to event recognition, the possibility must exist that this insight will somehow influence our manual segmentation of texts.

It would clearly be preferable for the entire manual segmentation process to be undertaken by subjects with no knowledge of the strategies employed by the system. However, the labour intensive and tedious nature of the process means that this is not a practical option.

Although we have tried hard not to let system knowledge affect manual segmentation, some form of external correlation is still necessary. The solution we arrived at was to have a group of naive subjects each segment the same corpus of eight texts, and then compare the segmentations both across subjects and with our own segmentations.

This raises a number of important issues. Whilst we want the subjects to segment the texts without knowing anything about the techniques used by CONTESS, we nevertheless want them to come up with a segmentation that uses the same notion of clauses and events. It is therefore important to issue the subjects with a clear set of guidelines that drive them towards the kind of segmentation we have in mind, without suggesting to them what explicit clues to look for in proposing, for example, an event shift.

We also need to make sure that the eight texts that we use constitute a representative sample of documents in the corpus. Although we considered simply selecting the texts at random, this was deemed unsatisfactory given the relatively small sample that we require and the wide range of document types in the corpus. Texts were therefore chosen on the basis of length (in terms of number of words), number of events, and document-type. The sample of texts used, together with a breakdown of the naive subjects' segmentations, can be found in appendix D.

Method

Having previously constructed an interactive text segmentation tool that uses a Web browser, it seemed natural to use this interface in the human correlation experiment. A series of linked HTML documents was therefore made available, and subjects were recruited through the use of general topic Usenet newsgroups. During the 24 hour period following the broadcast of the call for participation, over 150 individual segmented documents were received. No explicit identificatory information was requested from subjects; instead, IP[11] addresses and time stamps provide a workable (though technically not 100% reliable) method of subject distinction.

The information that test subjects were presented with consisted of three main sections. First, subjects were shown a page of instructions (reproduced within the appendix in section D.1), which briefed subjects *in general terms* on the text segmentation task. No mention was made in the instructions of the segmentation techniques used, either consciously by CONTESS, or unconsciously by the author. Instead, an example text was shown (the same eventful text as used throughout previous sections of this thesis) with event assignments marked up accordingly. Every effort was taken, therefore, to

[11] *Internet Protocol.* Each host on the Internet has a unique (though sometimes dynamically assigned) numeric IP address which can be used to identify the client from which a request is made.

ensure that subjects remained unaware of the strategies and heuristics being tested.

The second section of the experiment contained a set of eight texts that had been preprocessed to show clause, sentence and paragraph boundaries. The use of preprocessed texts imposes an artificial format on the text which subjects are forced to work with. Although in a few cases this undoubtedly results in distracting text formatting, it is necessary if we are to be able to compare text segmentations originating from different sources (human subject, program and author).

Subjects used the same text segmentation interface as used by the author and described in section 8.3.2 (also included in the process flow diagram in figure 8.14). Figure 8.15 shows a screenshot of one complete text as viewed with a popular web browser. The default state of the grid is shown as being minimally eventful. We believe that the adoption of this as a default is valid for several reasons; not only does it mean that grids with missing event assignments are not submitted, but it also reduces the effort required from subjects when segmenting (or rather, not segmenting) the longer and less eventful texts included in the test set (included in section D.2).

On submitting a segmented text, subjects were presented with an acknowledgement and an invitation to proceed to the following text in the corpus. On completing all eight documents, subjects received a page thanking them for their sterling effort and briefly describing the motivation behind the experiment, together with some hypertext links to relevant previously published work by the author.

In analysing the grids resulting from the subject-submitted texts, we are looking for an indication of the agreement between subjects in their segmentations. The *Kappa* statistic [Siegel & Castellan 88] (see also [Carletta 95] for a discussion of the relevance of this test for research in computational linguistics and cognitive science) can be viewed as a *coefficient of agreement for nominally scaled data.* Given that we can represent the subjects' grids as binary strings (with 1s and 0s being the nominal categories), this would appear to be a useful method of calculating the degree of agreement between subjects.

If we have N entities that can be assigned to one of m categories by k subjects, we can represent the assignments as shown in table 8.2 (adapted from [Siegel & Castellan 88]),

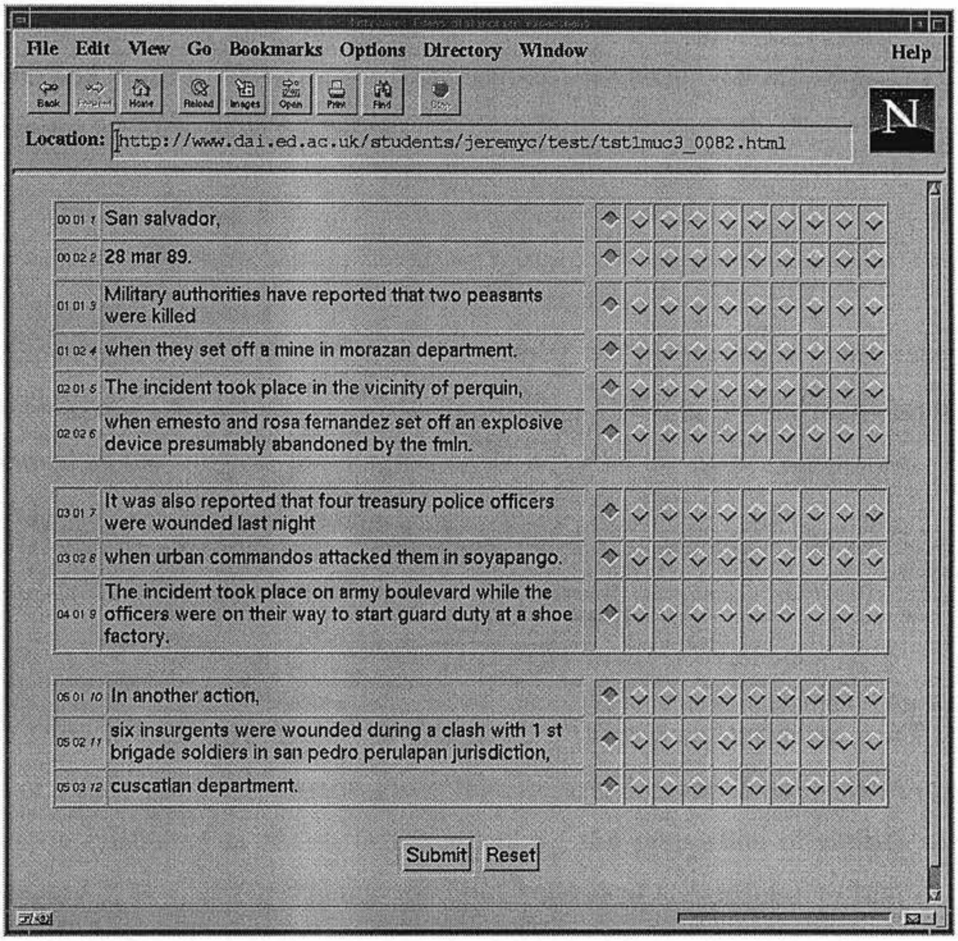


Figure 8.15: Screenshot of text as seen by correlation subjects

	category					
entity	**1**	**2**	$\cdots$	**j**	$\cdots$	**m**
1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1m}
2	n_{21}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\vdots$	$\vdots$			$\vdots$		$\vdots$
i	n_{i1}	$\cdots$	$\cdots$	n_{ij}	$\cdots$	n_{im}
$\vdots$	$\vdots$			$\vdots$		$\vdots$
N	n_{N1}	$\cdots$	$\cdots$	n_{Nj}	$\cdots$	n_{Nm}
	C_1	C_2	$\cdots$	C_j	$\cdots$	C_m

Table 8.2: Nominal category representation for κ statistic calculation

where n_{ij} is the number of times entity i has been assigned to category j. Assuming that every subject must assign every element to some category (which we do), each row will add up to k. The columns (representing categories) will add up to the sum of frequencies for each, as in 8.5 below.

$$C_j = \sum_{i=1}^{N} n_{ij} \tag{8.5}$$

The kappa coefficient κ is defined as the ratio of the proportion of times that the k subjects agree $P(A)$, adjusted for chance agreement $P(E)$, to the maximum proportion of times that the subjects *could* agree, again adjusted for chance agreement. This is shown in 8.6.

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{8.6}$$

When there is total agreement, $\kappa = 1$, and when there is no agreement $\kappa = 0$.[12] The value of $P(E)$, the proportion of times that we would expect k subjects to agree by chance, is calculated as shown below, where p_j, the proportion of entities assigned to category j, is $p_j = C_j/Nk$, i.e. the total number of assignments to that category divided by the product of the number of entities and the number of subjects.

$$P(E) = \sum_{j=1}^{m} p_j^2 \tag{8.7}$$

The total proportion of agreement, $P(A)$, is then calculated using the equation below.

$$P(A) = \left[\frac{1}{Nk(k-1)} \sum_{i=1}^{N} \sum_{j=1}^{m} n_{ij}^2\right] - \frac{1}{k-1} \tag{8.8}$$

Fitting our subjects' experimental data to the tabular format shown in table 8.2 above is simple. We have only two categories, 0s and 1s, and our entities are bits in a binary string. (Note that we would not want to have event numbers as categories and clause numbers as entities. As we argued in section 8.2, comparison of absolute

[12] Though, as we shall see in section 9.11, it is possible to have negative values of κ.

assignment positions is neither desirable nor particularly meaningful; rather, as we are interested in the relationships between assignments, binary strings offer us a more useful representation to evaluate.)

To calculate the κ coefficient for our subjects' segmentations, we first transformed the grid representations into binary strings. Having done this, we concatenated each subject's strings in turn into one long string, so that we were left with $k = 15$ long ($N = 3099$) binary strings.

As we pointed out in section 8.2.3, binary strings contain varying amounts of redundancy, with the exact amount dependent on the number of events the original grid describes. Bearing in mind Carletta's comments on the κ statistic [Carletta 95], i.e. that sensible choice of units is crucial in order to achieve realistic values for chance expected agreement, we repeated our experiment using binary strings with no inherent redundancy. These strings were formed by comparing in clause sequence each adjacent pair of clause-event assignments on a grid and, where an event discrepancy exists, this is denoted in the string by a `1`. A `0` denotes no event discrepancy. This reduced string formation technique is illustrated in figure 8.16.

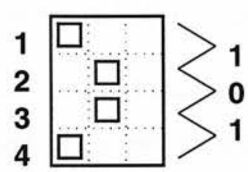


Figure 8.16: Reduced binary strings based on adjacency

Although these reduced strings lack the redundancy that is contained in the longer strings, they also fail to take into account long distance event relationships between clauses. Therefore, whilst the reduced string format is equally well suited to representing *boundaries* between events, it does not capture the element of *clustering* that exists between clauses.

Results of the naive subject correlation experiment are contained in section 9.6, with discussion of the results following in section 9.11.

8.4 Rule evaluation in the TAM and LAM

Up to now we have been discussing evaluation of CONTESS at two main levels: at the *system* level, i.e. as a complete text segmentation system; and at the *module* (or heuristic) level, where we are able to identify the contributions made by specific components of the system, and thereby make assumptions about the reliability of particular linguistic phenomena for text segmentation.

Both of these levels of evaluation treat the analysis modules as black boxes. Clearly, it would be useful to know in more detail, for example, which rules in the TAM and LAM tend to contribute to correct grid structures (when compared to manually constructed grids) and which to bad structures. Here we describe a mechanism we have implemented for this purpose.

For each text in the corpus, three information sources are used: a binary string representation of the manually segmented grid; an ordered set of binary strings (one for each configuration) of the computer segmented grids; and an ordered set of rules (ordered by clause number) used by the analysis module under investigation. This latter information source can be output by the TAM or LAM.

The evaluation mechanism consists of the following stages, and is applied (for either the TAM or the LAM independently) to each text in the corpus. For each computer generated binary string, compare every bit in the string with the corresponding bit in the manually generated string. If the hypothesis string is generated by a configuration that has the module (TAM or LAM) activated and the other analysis modules deactivated, *and* the two clauses represented by the current bit both contain a semantic representation generated by a rule in the module, then if the bits match (or do not match)[13], the score for both rules is incremented (or decremented) by one over the square of the length of the string (for purposes of normalisation). If, on the other hand, the hypothesis string is generated by a configuration that has all analysis modules *deselected* and the bits do *not* match, then the score for both rules that would have been associated with these bits is incremented by the same amount. (If the module is deselected and the bits *do* match, then no action is taken.) By positively scoring rules

[13] Matching is defined using the same negated *xor* operator as used in equation 8.1.

when their respective modules are deselected, we aim to further highlight the relative merits of those rules. The algorithm is represented formally in a short Perl program in appendix F, together with an example of the algorithm applied to the example text.

There are a couple of points to make about this algorithm. Firstly, rules are always scored in pairs (though the two rules may be the same). It does not make sense to score rules in isolation, as it is a defining property of semantic incompatibility (which results in constraint proposal) that two clauses (and hence two rules) are involved. Secondly, there are two ways in which a rule's score may be incremented. One way is for it to be jointly (and only potentially) responsible for a constraint that resulted in a correct assignment relationship (as defined by the manually segmented grid) while the module within which the rule resides is activated. The other is for it to be in an *incorrect* relationship while its own module (and the others) are *deactivated*, i.e. when neither the rule itself nor those of any other module could have contributed to the (now flat) grid structure.

The effect of this is that rules consistently (the procedure is applied to every text in the corpus) motivating correct grid relationships are rewarded, whereas rules consistently motivating incorrect relationships are penalised. This a rather crude evaluation strategy, which relies on a large number of documents to be meaningful. However, when used simply as a tool for highlighting potentially bad (and good) rules for manual consideration, it appears to be useful. Results of this finer-grained evaluation are presented in the Results chapter in section 9.3.

8.5 Other discourse segmentation evaluations

Given the limited amount of previous direct research on event recognition, it is unsurprising that there is even less work on evaluation of the techniques used. Event recognition programs have typically been integrated into complex information extraction systems, with the result that whilst the IE system as a whole may be quite rigorously evaluated, the evaluation generally treats the IE system as a black box. Besides making it very difficult to identify strong and weak event recognition approaches, this also means that there is little or no direct precedence for modeling our evaluation on.

8.5.1 General Electric's DPM

The closest we have found to an evaluation of an event recognition system is the analysis done by Iwańska [Iwańska 91] of the Discourse Processing Module (DPM, see section 2.4.2 for an overview), a program that was incorporated into the highly successful GE NLToolset system in MUC-3 and MUC-4 [Krupka *et al.* 91],[Krupka *et al.* 92].

DPM is functionally similar to CONTESS, but with one difference — a difference that is potentially very important for evaluation purposes. CONTESS has no concept of event relevancy, i.e. all events that it finds are deemed to be relevant. DPM identifies (and distinguishes between) relevant events, using the same notion of relevancy as that detailed in the MUC-3 guidelines (see section 1.2 of the thesis for details).

Although DPM is said to exhibit "good performance", the evaluation remains at a very informal level. The program was tested on a body of 100 previously unseen documents from the MUC-3 corpus, outputting for each document the sections of the original text that were deemed relevant, a percentage value showing the proportion of the original text that these sections constitute, and the number of distinct events (referred to as "segments") that were found. Examples of good, mediocre and poor performance are presented, which provide useful illustrations. As the evaluation is done entirely in isolation, DPM's approach cannot be quantitatively compared to other (human or automatic) segmentation techniques.

It would seem to us, however, that DPM is well suited to formal evaluation using the templates from the MUC-3/4 development corpus. Assuming that extracting time and location information from the individual segments is a relatively simple process, as we believe it would be, some degree of formal evaluation could be performed by correlating this information with the appropriate fields from the templates.

8.6 Summary

This chapter has presented the issues involved in the various forms of evaluation that we are undertaking in this thesis. We have seen how it is possible to evaluate:

- hypothesis grids with respect to a model grid;

- CONTESS's ability to generate good hypothesis grids;
- the author's ability to manually segment texts;
- the usefulness of types of natural language fragment in proposing grids;
- the contribution of individual clustering heuristics.

We have shown how rich descriptions of grids can be constructed, and discussed the advantages and disadvantages of the resulting binary strings. This has been illustrated with some example grid evaluations. We have also argued that more traditional evaluation metrics are unsuitable for the structures we are interested in. The κ statistic for agreement test was introduced, together with a description of how it can be applied to the naively segmented texts. Finally, we presented a brief description of the evaluation techniques used by other researchers in the area of event recognition.

The drawbacks of manual text segmentation have been pointed out. The labour intensive nature of this process provides further evidence for the need for more richly marked up texts in the information extraction field, a call that has already been made by those involved in the Tipster evaluations.

The next chapter shows what happens when we apply these evaluation techniques to CONTESS.

Chapter 9

Results and discussion

9.1 Introduction

Earlier chapters described the analysis modules and event manager that constitute the CONTESS system, and chapter 8 introduced a technique for quantitatively comparing grids. In this section we present the results of applying the full range of configurations of CONTESS to both the development corpus and the test corpus.

As we stated earlier in section 8.3.1, evaluation of our system consists of generating grids for each document in a corpus in each configuration. Consequently, for a body of 100 documents, and 128 different configurations of CONTESS, we obtain a table of results with 12,800 entries. Before presenting the results of the evaluation, three approaches to summarising this data (introduced in section 8.3.1) are illustrated using the now familiar example text.

9.1.1 Documents

The first of the three approaches involves listing, for each document in the corpus, the scores achieved by the ten configurations that, on average, scored the highest across all documents. As our example corpus contains only one document, table 9.1 contains just the results from the single document. (Tables with more documents below are split across several subtables.) The columns of the table represent, from left to right, the document number; the minimum and maximum scores achieved by any configuration for that document; the scores of the ten highest scoring configurations for the corpus;

the score achieved by a particular configuration selected for observation by the user (not used in this example); and the nature of the highest scoring configuration for that document *if* it is not one of the ten highest scoring for the corpus. The final two columns contain respectively the scores obtained by comparing a minimally eventful and a random grid of appropriate length with the manually segmented grid for that document. Below this main table is a key showing the identity of the top ten scoring configurations.

DOCUMENTS /hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
00	34.5	100.0	100.0	100.0	100.0	100.0	92.7	92.7	87.3	87.3	87.3	87.3			34.5	61.8
AVE	34.5	100.0	100.0	100.0	100.0	100.0	92.7	92.7	87.3	87.3	87.3	87.3			34.5	61.8

KEY:
1: s1101111 2: s1101100 3: s1101110 4: s1101101 5: s1100111 6: s1100101 7: s1001101
8: s1001111 9: s1101001 10: s1101011 C:

Table 9.1: Numeric document analysis of example corpus

As should be expected from a document whose purpose is to demonstrate, as far as possible, the analytical capabilities of CONTESS, four configurations manage to achieve a perfect score of 100, whilst the minimum score for any configuration was 34.5. As we can also see that a minimally eventful hypothesis grid scores 34.5, we can deduce that the lowest scoring configurations have all the analysis modules deselected — analysis modules being the only components in CONTESS that can propose an event distinction (and therefore a non-minimally eventful grid). This can be confirmed by looking at the next set of tables.

9.1.2 Configurations

Whereas table 9.1 summarised results in terms of documents, the third table[1] in section C.1 presents the results in terms of configurations, with the highest mean scoring configuration at the top and the lowest at the bottom. The columns correspond to: the rank of the configuration (where low numbers indicate a high ranking); the format of

[1] Out of respect for the reader, large results tables have been restricted to appendix C.

that configuration; the minimum and maximum[2] scores achieved by that configuration throughout the corpus, and the variance between the minimum and maximum; the mean score achieved, $\bar{X}$; the number of documents for which that configuration scored between 0 and 9, between 10 and 19 and so on, up to the number that scored exactly 100; finally, the standard deviation of the distribution for that configuration is displayed. Because there are 128 configurations, this table is split into smaller sub-tables.

As can be seen from this table, the top scoring configurations make use of the constraints provided by the TAM and LAM. In fact, the top four all achieve maximum scores. The lowest scoring configurations in the table either make use of no analysis modules or no gelling/clustering heuristics. This method of summarising the data therefore allows us to see in detail how particular configurations perform.

It would seem from looking at the configurations tables for the example text that certain modules, i.e. the TAM and the LAM, are instrumental in attaining high scores, whereas the CPAM is less useful. Of course, as we are dealing with only one document, we can only say that, *for this particular document*, this seems to be the case. As it happens, the only clause containing a cue phrase in our example document also contains a temporal phrase; furthermore, the sentence in which it is situated (repeated below) also contains a locative phrase. This means that the sets of constraints produced by the TAM and the CPAM overlap, and that, when clause gelling and clustering heuristics are taken into account, event shifts proposed by the LAM and the CPAM are in places identical.

- In a similar incident ten days ago, the JPF attacked an army installation in the town of Rivera.

The configurations table gives us a broad summary of the performance of individual configurations across a body of documents. It is also possible in glancing down the table to get some idea of how useful individual modules and heuristics are in building high scoring grids. As the establishment of these properties is one of the primary aims of this thesis, however, we need to know more accurately the contributions made by each module and heuristic. The next table allows us to do just this.

[2] It is extremely likely that, for any given corpus of substantial size, every configuration will have a maximum attained score of 100. All this means is that there is at least one document in the corpus on which all configurations scored 100 — for example, a minimally eventful document on which the analysis modules failed to be caught out by any spurious phrases.

9.1.3 Components

sT------	9.999987	sT------	2.718750
s-L-----	12.727263	s-L-----	3.718750
s--S----	-9.204544	s--S----	-2.468750
s---P---	12.272725	s---P---	3.656250
s----C--	9.659087	s----C--	2.718750
s-----F-	0.000000	s-----F-	0.000000
s------U	16.022706	s------U	5.156250

Table 9.2: Numeric (left) and ranked (right) component analysis of example corpus

Table 9.2 represents the contribution made by each module and heuristic in constructing high scoring grids. The values for each are obtained by selecting a particular aspect of the configuration and toggling its selection between on and off while varying the remaining aspects of the configuration in every possible permutation, and keeping a total of the scores in each state. Total net gain (or loss) then results from adding up the differences between the scores produced by each pair of complimentary configurations, and is shown alongside each configuration in table 9.2. Identical values may be obtained through multiple regression analysis, with the configuration score as the dependent variable and the states of the seven components as independent variables.

As the example text was written for the express purpose of concisely demonstrating the features of CONTESS, it is hardly surprising that the results show most of the modules and heuristics contributing so positively to the segmentation process. The figures in table 9.2 should therefore not be taken seriously. The only figure of any real interest in the above is for the term vector frequency analysis component, which did not contribute at all (either positively *or* negatively) to the grid construction process. Frequency analysis is only very rarely called into use — although it is the first of the gelling techniques to be considered for unconstrained clauses, the threshold above which a clause must score is such that, in effect, assignment positions are rarely decided on the basis of frequency analysis. (See further discussion in section 9.9.2 below).

9.1.4 Rank analysis

As the absolute value of raw scores is not clear, and, in particular, averaging raw scores is technically meaningless, we also present the three analysis approaches described above in terms of *ranks*. This section gives a rank analysis example of each table type seen in the previous three sections. Table 9.3 shows the example corpus (with its single document) with configurations ordered by rank rather than raw score. Again, low numbers indicate a high ranking.

<table>
<tr><td colspan="14">DOCUMENTS
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile.ranked</td></tr>
<tr><td></td><td colspan="2"></td><td colspan="11">OVERALL RANKED CONFIGURATIONS</td></tr>
<tr><td>DOC</td><td>MIN</td><td>MAX</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>(C)</td><td>sTLSPCFU/Q</td><td>[||]</td><td>[?]</td></tr>
<tr><td>00</td><td>1.0</td><td>18.0</td><td>1.0</td><td>1.0</td><td>1.0</td><td>1.0</td><td>2.0</td><td>2.0</td><td>3.0</td><td>3.0</td><td>3.0</td><td>3.0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>AVE</td><td>-</td><td>-</td><td>1.0</td><td>1.0</td><td>1.0</td><td>1.0</td><td>2.0</td><td>2.0</td><td>3.0</td><td>3.0</td><td>3.0</td><td>3.0</td><td></td><td></td><td>-</td><td>-</td></tr>
<tr><td colspan="17">KEY:
1: s1101111 2: s1101100 3: s1101110 4: s1101101 5: s1100111 6: s1100101 7: s1001101
8: s1001111 9: s1101001 10: s1101011 C:</td></tr>
</table>

Table 9.3: Ranked document analysis of example corpus

Consequently, rather than seeing the ten configurations that had the highest absolute scores (even though these ten actually only share three scores), we are shown *an example* of a configuration that is responsible for each of the top ten scores.[3] The "overall ranked configurations" section of the table shows where the overall first, second etc. configurations (i.e. the example configuration selected and presented at the bottom of the table) ranked for each particular document. Tied ranks are used; consequently, if all 128 configurations score full marks, they will all share ranks of 64.5.

Similarly, the latter tables in section C.1 list the configurations in order of rank, but show the distribution of ranks rather than the distribution of scores — that is, how many times each configuration achieved the highest rank (*not* ranked first), how many times the second highest, and so on up to "more than tenth". The "minimum" and "maximum" columns are replaced with "top" (actual top rank achieved in the corpus) and "btm" (bottom rank achieved in the corpus). Table 9.2 also presents a component analysis performed using ranks.

[3] Clearly there are likely to be multiple configurations for each score — for example, four configurations score 100% in this corpus.

Results

The results component of this chapter is divided into four sections, each based on one major evaluation. These evaluations include: an initial evaluation using the development corpus; a second evaluation using the same corpus but following TAM and LAM rule performance analysis; a third evaluation with the CRaP document filter activated; and a fourth evaluation using an unseen test corpus.

9.2 Development corpus results I

This section presents the results of CONTESS applied to the development corpus. To maintain readability, only component analyses are included here — the complete set of results tables for this section may be found in the appendix in section C.2.

sT------	1.111358	sT------	5.199063
s-L-----	-1.733543	s-L-----	-6.810312
s--S----	0.026074	s--S----	0.140469
s---P---	-0.247831	s---P---	-1.057969
s----C--	1.660815	s----C--	4.437187
s-----F-	-1.900357	s-----F-	-5.032813
s------U	0.804530	s------U	1.460938

Table 9.4: Numeric (left) and ranked (right) component analysis of development corpus using original grammars

As can be seen in table 9.4, the `T`, `C` and `U` components (that is, the TAM, clause gelling heuristic and CPAM) contribute positively, and significantly so ($p < .01$). Sentence gelling, `P`, is marginally negative, and the LAM, `L`, is marginally negative in its contribution to configuration scores. Neither of these latter two results, nor the clustering strategies, `S`, or frequency analysis, `F`, are statistically significant for values of $p < .05$.

The configuration tables in section C.2 reflect this component analysis. Configurations with the TAM, sentence gelling, clause gelling and CPAM selected are ranked towards the top of the table, whereas configurations without these components are ranked much lower. Note also that the standard deviation of lower scoring configurations increases as we progress down the ranks, indicating a wider spread of scores — and, specifically, a lower minimum bound.

9.3 Development corpus results II

In this section we show the results of running CONTESS on the development corpus after the application of the rule evaluation technique introduced in section 8.4. Rule evaluation identified and scored rules in the TAM and LAM grammars as shown in table 9.5.

rule	score
R1	0.005384
R2	0.004656
R3	0.000145
R4	0.001792
R5	0.000584
R6	0.020915
R7	0.002527
R9	0.003737
R1i	0.015031
R2i	0.000142
R10	0.001234
R11	0.003259
R20	0.000554
R12	0.001675
R21	0.000989
R1ii	0.011063
R2ii	0.002202
R22	0.000690
R23	0.021312
R16	0.001595
R18	0.000506
R19	0.001746
R12i	0.007742

rule	score
R1	-0.001256
R11	0.001686
R2	0.005130
R99	0.002393
R3	-0.000525
R12	0.000489
R4	0.006003
R5	0.001975
R6	0.000375

Table 9.5: Rule evaluation of original TAM (left) and LAM (right) grammars

Although we are unwilling to infer much from the precise scores attributed to each rule, there does at least appear to be a correlation between high scores and those rules that we intuitively feel are both widely and successfully applied. Investigation of the use of rules that were attributed negative scores (that is, rules R1 and R3 in the

LAM) led to the decision to eliminate these rules from the LAM grammar. Section 9.10 below contains a discussion of the reasons for the poor performance of these eliminated rules. The results of repeating the development corpus segmentation with the modified grammar are shown in table 9.6.

sT------	1.262345	sT------	5.606563
s-L-----	-1.426397	s-L-----	-6.387813
s--S----	0.026936	s--S----	0.134219
s---P---	-0.222732	s---P---	-1.016094
s----C--	1.765016	s----C--	4.349687
s-----F-	-1.809405	s-----F-	-4.894063
s------U	0.807474	s------U	1.428437

Table 9.6: Numeric (left) and ranked (right) component analysis of development corpus using modified grammars

As is clear from the comparative bar chart in figure 9.1, rule evaluation results in increased scores for all components. Given that only the LAM grammar was altered, it is to be expected that the largest differences will be in the scores of this component, with score variations among further components a result of component interaction.

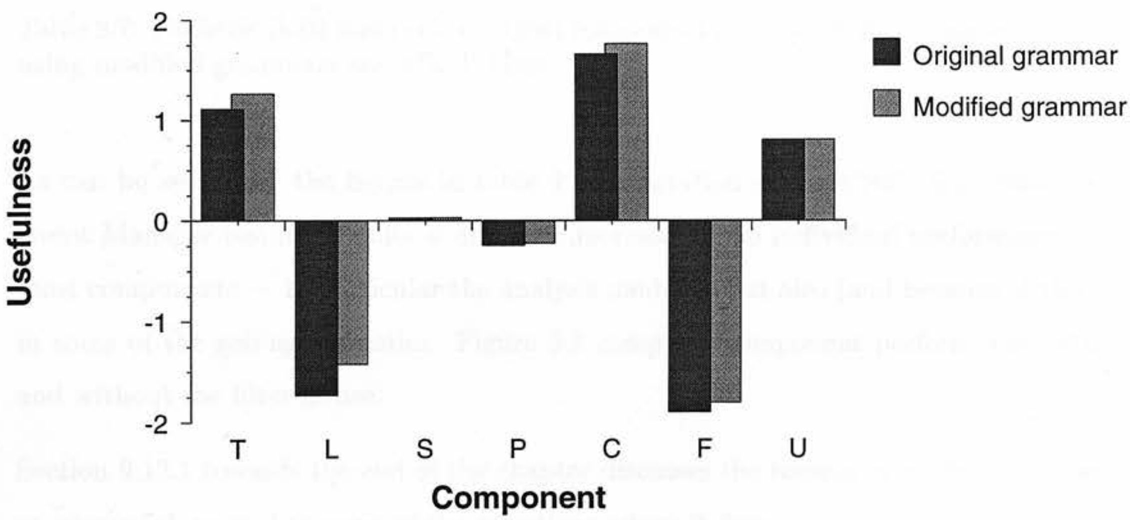


Figure 9.1: Comparison of numeric component analysis results for original and modified grammars

It would appear, then, that the elimination of rules with negative rule evaluation scores results in a positive change in the score of the relevant configuration component. To further validate this hypothesis, we created a third grammar for each of the TAM and the LAM containing the same rules as in the previously modified grammar, but with rules assigned scores that were positive yet less than 0.001 also eliminated. Component analysis revealed that the results, although marginally better than with the original grammars, were significantly worse than when only negative scoring rules were eliminated.

9.4 Development corpus results III

This section presents the results of activating the CRaP filter (described in section 7.4) in a third preliminary evaluation using the development corpus. The modified grammars are maintained from the previous section.

sT------	2.347637	sT------	9.084062
s-L-----	-0.085065	s-L-----	-2.484062
s--S----	0.015399	s--S----	0.107969
s---P---	0.209383	s---P---	0.173906
s----C--	2.636099	s----C--	5.892188
s-----F-	-1.541117	s-----F-	-4.314063
s------U	1.117817	s------U	2.578438

Table 9.7: Numeric (left) and ranked (right) component analysis of development corpus using modified grammars and CRaP filter

As can be seen from the figures in table 9.7, integration of the CRaP filter into the Event Manager results in quite a dramatic increase in the individual performance of most components — in particular the analysis modules, but also (and because of this) in some of the gelling heuristics. Figure 9.2 compares component performance with and without the filter in use.

Section 9.12.1 towards the end of the chapter discusses the reasons why the filter was so successful, as well as some of the situations where it fails.

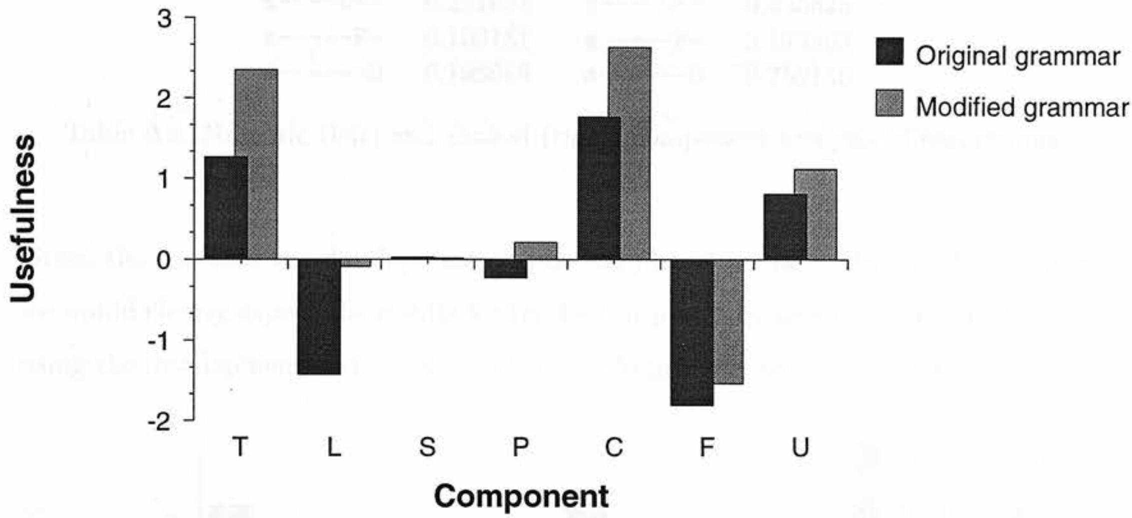


Figure 9.2: Comparison of numeric component analysis results with CRaP filter activated/deactivated

9.5 Test corpus results

This section contains the results of segmenting the 100 documents in the test corpus. The results in each of the three analysis formats — documents, configurations and components — are in appendix C.5, with only the shorter component analysis tables included here. For this final evaluation, both the modified grammar and the CRaP filter are activated.

sT------	1.361840	sT------	4.652500
s-L-----	-2.691750	s-L-----	-8.802188
s--S----	0.008032	s--S----	-0.041562
s---P---	0.006010	s---P---	-0.091875
s----C--	0.291661	s----C--	0.695625
s-----F-	0.100181	s-----F-	-0.136563
s------U	0.196969	s------U	0.750156

Table 9.8: Numeric (left) and ranked (right) component analysis of test corpus

Given the function the development corpus has played in the evolution of CONTESS, we would clearly expect the results for the test corpus to be lower than those obtained using the development corpus. As can be seen from table 9.8, this is indeed the case.

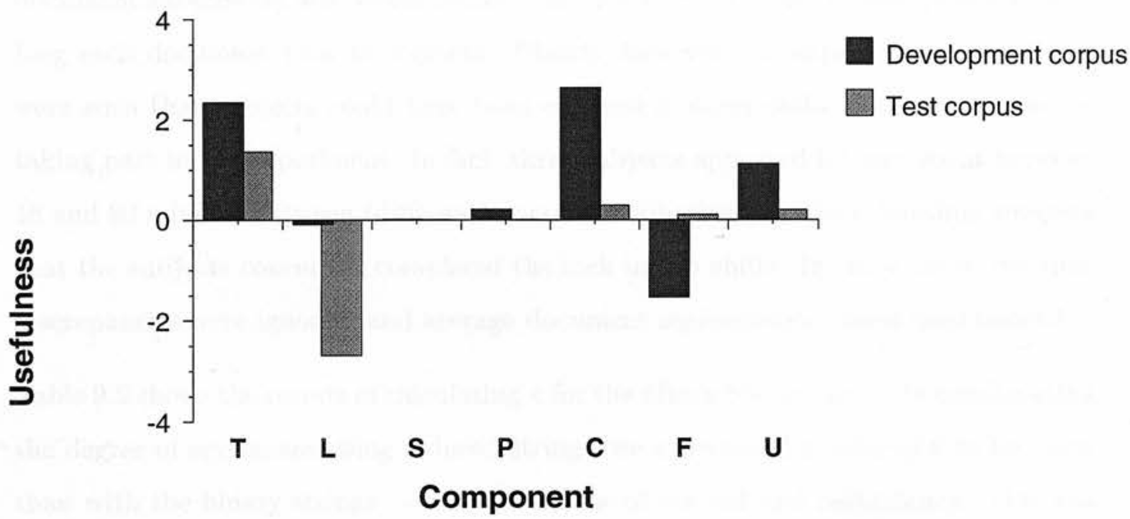


Figure 9.3: Comparison of numeric component analysis results for development and test corpora

$P(A)$=.781656
$P(E)$=.500396
κ =.562966

Table 9.9: Values of κ for test subjects using binary strings

$P(A)$=.870687
$P(E)$=.726035
κ =.527993

Table 9.10: Values of κ for test subjects using reduced binary strings

9.6 Human correlation results

A complete set of eight segmented documents was received from fifteen different subjects; 44 individual documents were submitted by another fourteen subjects who started, but did not complete, the segmentation task, and these were ignored. Only complete sets of documents were considered.

The average time taken by each subject was approximately 2 minutes per document — nearly 17 minutes for the complete set, which was as we had expected. As each document submission was automatically time-stamped, we have a useful guide to how long each document took to segment. Clearly, however, the experimental conditions were such that subjects could have been engaged in other tasks at the same time as taking part in the experiment. In fact, three subjects appeared to have spent between 15 and 90 minutes between (different) document submissions, which certainly suggests that the subjects concerned completed the task in two shifts. In these cases, the time discrepancies were ignored, and average document segmentation times used instead.

Table 9.9 shows the results of calculating κ for the fifteen test subjects. In recalculating the degree of agreement using reduced strings, we expected the value of κ to be lower than with the binary strings — simply because of the reduced redundancy. This was indeed the case, as shown in table 9.10. However, the drop in the value of κ was only very slight at $-.035$, removing any doubts we had about the effect of string redundancy on agreement measures.

The above agreement tests were performed using only the naive subjects. Consequently, although the results suggest that subjects agree amongst themselves, it says nothing about whether they agree with the "expert coder" — i.e. the author. The above two agreement tests were therefore repeated with the author's segmentations included in the pool of $k = 16$ subjects.

	expert coder? yes	no
full strings	.559378	.562966
reduced strings	.537749	.527993

Table 9.11: Comparative representation of κ values

Table 9.11 summarises the values of κ for each of the four conditions — full/reduced strings, expert coder included/not included. As can be seen, the inclusion of the expert coder in the subject pool results in marginally (and nonsignificantly) lower agreement when using full strings, but higher agreement when using reduced strings, a result that we are unable to explain.

Discussion

The remainder of this chapter presents a discussion of the above results, both in terms of inter-evaluation comparisons (i.e. how components fared across corpora[4] and test conditions) and intra-evaluation comparisons (i.e. how components performed relative to other components within corpora and test conditions). The discussion begins with a general overview of the performance of CONTESS, and continues by covering the analysis modules, the gelling heuristics used by the Event Manager, rule evaluation, human correlation and further issues arising from evaluation of CONTESS.

9.7 General performance

Evaluation of CONTESS using both the development corpus and the test corpus showed that high scoring configurations were indeed achieving higher scores than would be obtained using either a random segmentation approach or a minimally (or maximally) eventful one. From this we can make a number of inferences, both about the nature of the documents in the corpus (see section 9.12.3 towards the end of this chapter) and about the usefulness of CONTESS as a document segmentation tool.

9.8 Analysis modules

The relative performance of the three analysis modules (TAM, LAM and CPAM) varied widely. The TAM, easily the most successful component in the development corpus, also ranked highest amongst the components in the test corpus. The LAM, which in itself contributed negligibly to the performance of CONTESS on the development corpus, resulted in a significant overall *decrease* in performance on the test corpus. Further reflecting the expected drop in performance between development and test corpora, the CPAM experienced a decrease roughly proportional to that of the TAM.

[4] The example corpus is excluded from this discussion.

9.8.1 Temporal analysis

The general methods used by the TAM — that is, the island-driven shallow parsing, the concentration on punctual rather than durative temporal phrases, the typology adopted, and the nature of the interpretations given to the parsed fragments — would appear to have met with some success, as is shown by the consistent performance of the TAM on both the development and test corpus.

As can be seen in the results tables earlier in this chapter, the TAM was clearly the most successful component in CONTESS. Here we shall look at some of the cases where it performed well, and also at cases where it performed less well.

Using the same rule evaluation techniques as described previously in sections 8.4 and 9.3, it is possible to get some insight into the *relative* (as opposed to absolute) performance of rules within the TAM and LAM. It is important to bear in mind, however, that the success of the TAM (and LAM) is not necessarily directly linked to the success (in terms of frequency of usage and correctness of interpretation) with which the rules within the modules' grammars are applied. That is, a rule might be frequently used, and result in semantic representations that are exactly as we had intended them to be, without contributing much to the overall success of the analysis module. It is conceivable that, although the class of phrases is accurately identified, it simply is not used to assign punctual time stamps to events in the text. By combining the rule evaluation scores with information detailing the frequency and position of rule application in individual documents, we should, however, be able to identify any such cases.

The most successfully applied rule within the TAM in terms of contribution is rule R11, which is responsible for date phrases such as 'june 30' and '21 nov 95'. These temporal phrases were extremely common, especially in the headers of documents, with the vast majority of documents in both the development and test corpora containing an instance in the header. They were less common in the body of documents, though still quite frequently used.

TAM grammar rule R11 does not require that a preposition precede the date expression. Although the stylised nature of the header section of news documents means that a preposition such as *on* is redundant, even in the main body of the document prepositions are often omitted — unless the intended meaning is *before* or *after* a particular date. This seems to be particularly common in American English in many types of temporal phrase where the intended meaning is *on* or *at*.

- the first attack was launched *march 21st*...
- the president arrived in san salvador *friday*...

Further examples, such as 'the 19 march elections', are also correctly identified. However, a small subset of phrases in which temporal expressions were used as part of a group proper name, e.g. 'the 19 april movement', were incorrectly identified as temporal expressions in their own right. Such cases were rare.

A second highly successful rule within the TAM (in both the development and test corpus evaluation) was rule R16, which assigns interpretations to phrases such as 'today', 'yesterday' and 'sunday at 4.15'. Examples of applications of this rule include the following.

- it was officially reported that a policeman was wounded *today* when urban guerrillas attacked the guards at a power substation located in downtown san salvador.
- the cerezo agenda included a dinner he will host *tonight* for his counterpart carlos andres perez.
- *today*'s attack is the 25th during 1989.

Rule R19 was also frequently used, though more so in the development corpus than in the test corpus. The rule covers phrases such as 'on friday', 'at nightfall' and 'in january', with optional date phrases as covered by rule R11 (above) following. Examples of the use of this rule include:

- garcia alvarado, father of six, was appointed attorney general *on 23 december 1988*.
- *on 1 november* urban guerrillas attacked the headquarters of the 1st infantry brigade with rockets.
- attention: *during yesterday*'s attack, with people's artillery on 4th infantry brigade troops, we inflicted eight casualties.

A potential problem that we encountered with temporal phrases of this form was the use of *since* and *until*. Because our temporal representation formalism has no way of representing durations, only points, we cannot fully interpret phrases such as 'since january 1989'. However, this is not as significant a problem as it might at first seem. Given that the kinds of temporal phrases that indicate distinct events in a text nearly always involve a punctual reference of some sort (at whatever level of granularity), open-ended (either start- or end-point) durations of this form are not of much interest to us.

However, although we have no need of an interpretation of phrases modified by *since* or *until*, we cannot simply ignore the presence of such phrases. Because the TAM will always attempt to find the largest phrase possible, simply omitting *since* from the grammar will result in the phrase 'january 1989' (using the example from the previous paragraph). This is clearly not what is intended by the original temporal phrase, and has a quite different meaning.

Consequently, our solution is to include the words *since* and *until* in all versions of the grammar[5], but to assign a "null interpretation" to phrases containing them — in this way, they receive an interpretation (thus preventing the parser from attempting to find an interpretable subphrase), albeit an interpretation that will be compatible with any other interpretation.

This might seem to be an unsatisfactory solution. As a result of this decision, the interpretations of phrases such as 'since 1980' and 'in 1973' would be compatible. However, bearing in mind what we are trying to achieve in interpreting temporal phrases, i.e. distinction between multiple events, this need not be a serious problem. In such a situation, the TAM would not propose that the pair of clauses containing these phrases are event-coreferential — this is not the role of the TAM. Rather, there would simply be an absence of a constraint between the two clauses; and, given that it is unclear whether durative phrases can be seen as indicating new events, this is precisely what we want.

Another grammar rule that proved useful in the TAM was rule R10, which is responsible for phrases of the form 'at 0030 today', 'at daybreak on 16 november' and 'at night', i.e. temporal expressions describing a point during a (not necessarily explicit) day.

[5] as *durative prepositions*

Examples from the test corpus include the following.

- however, *at approximately 1700 today* a bomb exploded inside a fast food restaurant.
- aguilar marroquin was arrested *at 1500 on 9 december* after having participated in an attack on an urban public bus near juan pablo ii boulevard.
- *at noon today*, clashes with 1st military detachment troops was reported in la cruz hill, san miguel de mercedes jurisdiction. we repeat: *today at noon*, fighting was reported in la cruz hill, san miguel de mercedes, chalatenango department.

A less successful rule was R20, which is designed to handle phrases such as 'last november' and 'this january' (i.e. *the January of this year*). However, whilst the first example below is satisfactory, the second two most clearly are not.

- aguilar marroquin joined the erp *last march* and...
- anything *that may* lead to violence, whether it is street disturbances or bombs, should not be permitted...
- they noted that *this may* have been a terrorist attack because...

We have presented the above examples, and all others, in single case, as this is the form in which CONTESS encountered the texts. Consequently, there was no way to distinguish between *that may* and *that May.* With mixed case text now becoming prevalent in newswires, such distinctions are easily made.

9.8.2 Locative analysis

In this section we shall attempt to identify the reasons behind the poor contribution made by the LAM in both the development and the test corpus.

One possible area at fault is the nature of the rules within the LAM. To see whether this is indeed the case, we can use rule evaluation techniques, in exactly the same way as with the TAM earlier, to identify the performance of individual rules within the LAM.

There are relatively few top-level rules in the LAM (eight in the modified grammar compared with seventeen in the modified TAM), yet even so two of the rules appear to dominate in terms of frequency and success of application. These are rule R12, which is used to identify references to habitation types, e.g. 'the town' and 'the village' (the motivation behind this being that a single location is unlikely to be described using two different habitation types), and rule R99, which relies on place names marked up during preprocessing, e.g. in `place#[la,paz]`.

Whilst rule R12 was widely applied, it was rarely instrumental in proposing constraints between clauses — only in one document did it identify two clauses that were deemed to be incompatible, one referring to

- the peasant *village*

and the other to

- medellin, *a city* where the world's most powerful cartel has its base of operations

As it happened, these clauses *were* actually part of different events. Most of the applications of this rule, however, simply identified phrases such as 'the country' (which is not incompatible with any other geographical entity in the LAM hierarchy, see figure 5.1). As much of the time these phrases were not in clauses judged (by the author and CONTESS) to be in event constraint relationships, rule evaluation rewarded them accordingly. This shows that the rule evaluation scores cannot be taken simply as an indication of how good (or bad) the rules under consideration are — instead, they show the correlation between rule application frequency and the proportion of times that the clauses to which the rules were applied have been assigned to correct event relationships. If, as we believe is the case with rule R12, the phrases identified by a rule play an insignificant role in the signalling of event shifts, rule evaluation will nevertheless reward that rule, providing that it has a high frequency of application and results in a relatively low number of incorrect event relationships. Therefore, although on the surface it would seem that rule R12 is quite a successful rule, it is in fact simply an irrelevant rule.

Rule R99 has by far the highest application rate of any rule in the LAM. This is partly due to the presence of locative phrases in the document headers, in the context of which rule R99 does not require a preposition to interpret a marked up location. Analysis of the phrases identified by rule R99 presents a quite different picture to that seen with rule R12. Instead of producing multiple but compatible interpretations, rule R99 seems to have been responsible for generating multiple *incompatible* interpretations — i.e. it results in a large number of proposed constraints. There are many examples of this in the test corpus; in one particular document, rule R99 was applied eleven times, resulting in 59 constraints. As it happened, this example (included in appendix E.2) was particularly eventful, and was originally identified as such by CONTESS.[6] However, there are many other documents containing multiple locative phrases as described by rule R99 that were *not* judged by the author (nor by the naive coders — see section 9.11) to be eventful. One such document, which is reproduced in full in appendix E.3, describes a complex drugs smuggling operation, paying particular attention to the geographical aspects of the process. The following is a short representative extract.

the most important part of the trafficking, however, is carried out in paraguay and bolivia. gerson palermo is responsible for part of it, while the rest is shared by hundreds of bolivian traffickers. palermo obtains ether and acetone from formosa, argentina. he takes these chemicals aboard small planes to the towns of pilar or encarnacion, in paraguay. from there, another group takes the 200-liter drums along the chaco highway to puerto guarani, in bolivia. bolivian traffickers take care of the rest of the journey to santa cruz de la sierra. in return, palermo receives cocaine paste which is shipped to asuncion and, from there, to pedro juan caballero for distribution to the processing units. it is believed that the traffickers use some ranches near porto murtinho (mato grosso de sul), near the paraguayan town of bahia negra, to process large quantities of cocaine.

At a certain level, it is true that this document contains multiple events — it describes several different journeys, transactions and processes. However, at the level of events that we are interested in, it is not eventful. Instead, the activities that it describes are quite clearly stages in an elaborate process — and therefore at most sub-events within a larger "drug-smuggling" event. We believe that, in this way, the activities described above are analogous, for example, to the construction, priming, positioning and detonation of an explosive device; that is, the component activities of a single event. Such detailed descriptions are more relevant in describing drug smuggling simply because of the scale and complexity of such operations.

This therefore seems to point towards a more general inappropriateness of the LAM as it currently stands in the CONTESS framework. Without a script-like knowledge of events (or even any knowledge of event *types*, about which more is said in section 10.3), spatially distributed operations such as that described in the above document will cause

[6] However, this document also received a high CRaP score, and consequently with the CRaP filter activated, the LAM constraints were not taken into consideration. See section 9.12.1 for further discussion of the CRaP filter.

the LAM to propose spurious constraints. As single events are rarely described in a temporally distributed manner (possibly due to the typically instantaneous nature of the events that our approach is aimed at), this is not such a problem for the TAM. This suggests, we believe, that locative phrases in general are not useful indicators of multiple events.

9.8.3 Cue phrase analysis

We were quite surprised by the relatively high score that the CPAM received as a component during preliminary evaluation using the development corpus, particularly as the approach used was extremely naive — i.e. that the introduction of one of a small set of cue phrases signals the introduction of a new event in the discourse.

One factor partly responsible for the decreased contribution of the CPAM between the development and the test corpora is the difference in the frequency of cue phrases in the two corpora: the development corpus contains 40% more cue phrases than the test corpus. In both corpora, the phrases 'meanwhile' and 'in addition' constitute the bulk of those identified. Nevertheless, the fact that such a simple strategy continues to contribute positively when applied to the test corpus seems to suggest that the application of shallow text processing techniques to text segmentation tasks warrants further investigation.

9.9 Heuristics

The heuristic components of the Event Manager — the clustering strategies, frequency analysis, sentence- and clause-gelling — exhibited far less dramatic performance (positive or negative) than the analysis modules. This section attempts to explain the observed behaviour of these gelling components.

9.9.1 Clustering strategies

After initial evaluation with the development corpus, we did not expect the choice of clustering strategy to have a significant effect on system performance — simply because there are relatively few cases where altering the strategy results in clustering variation;

the choice of clustering strategy is only of any significance in the rare situation where *constrained* clauses have multiple possible positions on the grid.

As can be seen from the component bar chart in figure 9.3, the choice of clustering strategy adopted doesn't actually affect system performance on the test corpus significantly.

9.9.2 Term vector frequency analysis

The frequency analysis heuristic used by the Event Manager also failed to make a significant contribution to the segmentation process. By holding fixed the state of the rest of the components of CONTESS, we were able to calculate through a series of evaluations on the development corpus the optimum settings for the various parameters involved in frequency analysis.

The parameters we modified included the units to analyse (word stems or character N-grams of a specified size), the threshold to use for determining similarity between sentences in the output and the use of term weight normalisation. We found that the optimum configuration consisted of using character trigrams, with a threshold of 0.30 and normalisation turned off. Although the reason for not using normalisation was initially empirically motivated, it is not difficult to see why it is undesirable in the context in which it is being used.

The purpose of normalisation in the traditional Smart approach to text analysis (such as that we described in section 7.3.1) is to prevent longer documents (or in our case sentences), which generally contain more terms and higher term frequencies, from being preferred to shorter documents (i.e. sentences), with fewer terms and lower frequencies.

In theory this is a fair assumption. In practice, however, when working at the level of sentences (as opposed to whole documents), normalisation results in short sentences, with only a few terms, being assigned high similarity scores — on the basis often of only one or two words (after common and function words have been removed). This finding reflects that of Salton and Buckley [Salton 71], who also point out that normalisation is not practical at low levels.[7] In summary, the rejection of normalisation

[7] Hearst's use of *pseudosentences* in her word-frequency based TextTiling algorithm (see section 2.4.5,

therefore favours longer sequences of matching terms over shorter ones, which, at the sub-document level, is in practice more useful.

In the case of the development and test corpus evaluations, the frequency analysis component has only a marginal effect on the overall segmentation process. The conclusion that we must draw from this is that we are simply applying the analysis technique at too low a level. Although we had ruled out operating at the clause-level in advance for this reason, it seems that, even without the use of normalisation, term vector frequency analysis techniques are not appropriate at the sentence level. This hypothesis is supported by the observation that frequency analysis gives better results when applied to longer documents — that is, documents with more terms and a wider spread of term frequencies.

9.9.3 Sentence and clause level gelling

All of the gelling heuristics used by the Event Manager are dependent on the performance of the analysis modules. Unless the analysis modules propose an event shift, the gelling heuristics have nothing to work with. A consequence of this is that the observed performance of the gelling heuristics depends to a certain extent on the performance of the analysis modules. If the analysis modules perform well, then the component scores of the gelling heuristics operating on the resulting grid assignments will be affected accordingly.

This is particularly true for the sentence and clause gelling components. As we saw in sections 7.3.2 and 7.3.3, the effect of these heuristics is to allow the EM to extend the suggestions made by the analysis modules further into the text, by gelling together surrounding sentences (in the case of the former heuristic) and clauses (in the case of the latter). The clause gelling heuristic in particular controls the position of a high proportion of the assignments of unconstrained clauses to the grid. This is because there are more clauses in the average sentence than there are sentences in the average paragraph — in both the development and the test corpus, the *theoretical* limit on the

where token sequences of fixed lengths are used in lieu of syntactically motivated structures, circumvents issues of normalisation. However, the resulting adoption of artificial sequence boundaries means that segmentation across natural (sentence/paragraph) boundaries is made harder.

number of clauses that could be assigned positions using the clause gelling heuristic is approximately 60%[8] of the total number of clauses in any document. For sentence gelling, the figure is slightly less than 30%.[9]

This component interdependency explains the relationship between the drop in performance observed in the analysis modules between evaluations and that observed in the gelling components — and, in particular, the performance of the clause gelling heuristic.

9.10 Rule evaluation

The rules eliminated as a result of the rule evaluation techniques described in section 8.4 exhibited a range of properties that made their removal from the LAM grammar desirable. LAM rules R1, which is responsible for identifying phrases such as 'department of ayacucho', and R3, which covers phrases such as 'comalapa airport', i.e. utilities preceded by place names, tended to be used in the corpus in a manner that did not locate events, but rather for describing routes and intended destinations — e.g. 'the road to la libertad port'. In one case where such a description was actually used to define an event location, more general location information (such as the city in which the utility is situated) was available. No documents in the corpus contained multiple events situated within the same city, but occurring at different utilities within that city (e.g. 'la paz airport' and 'the port of la paz'), which suggests that it is not necessary to be able to distinguish between locations at this level.

It is difficult to be specific about the reasons behind the observed increase in the TAM performance following deletion of the above LAM rules. One hypothesis is that module interaction is the primary cause, i.e. with the removal of these particular LAM rules the TAM is more in control (in terms of proposing constraints that are not mixed with incorrect LAM constraints). However, we believe that it is more likely that the reason lies in the way in which the components are scored; in scoring a component, its

[8] The theoretical limit is equal to the number of clauses in a document minus the number of sentences (assuming the one constrained clause per sentence needed to motivate the gelling).

[9] This figure represents the number of sentence-initial clauses in non paragraph-initial position (i.e. the number of non paragraph-initial sentences).

performance is examined in every configuration, with the component under consideration alternately toggled on or off. Because of this, the relative performance of each component will have a very slight effect on the score attributed to the others.

9.11 Human correlation

Establishing what constitutes significant values of κ is difficult. Siegel & Castellan describe a test for significance based on the premise that, for large values of N, κ is "approximately normally distributed with mean 0." The variance that they give seems somewhat arbitrary (see [Siegel & Castellan 88]). According to the example in their book, $\kappa = .41$ is significant at $p < .01$. On the other hand, Krippendorff [Krippendorff 80] claims that researchers in the field of content analysis typically view $\kappa > .8$ as indicating reliable agreement, with $.67 < \kappa < .8$ sufficient for "tentative conclusions" to be drawn. Clearly there is a significant — and we use the word loosely — difference in opinion here. Based on Siegel & Castellan's measure of significance, our result is extremely significant, due to a certain degree, no doubt, to our large N.[10]

Perhaps the easiest way to envisage the degree of agreement between the naive coders on individual documents is to use a graphical representation as contained in the appendix in section D. Here we use a "bird's eye" view of multiple grid interpretations of a document — see section D.3 for a complete description.

In some of the documents the agreement images show a remarkable degree of agreement between coders (including the expert). The images for documents 4 and 6 in particular are striking in their uniformity, and the κ values shown along side the images reflect this. One of the documents, number 3, has a *negative* κ value — contradicting Siegel & Castellan's claim that $0 \leq \kappa \leq 1$.[11] The document in question was included in the corpus specifically to see how naive coders responded to uneventful rhetoric, and did not contain, in the view of the author, multiple events. Ten of the fifteen naive coders agreed with the author in this respect; three identified two events in the text (two

[10] Research in information retrieval has shown that agreement between human coders (as to whether or not a document is relevant) could be as low as 60% [Cleverdon 91] — unfortunately, we do not have any indication of what this equates to in terms of the κ coefficient.

[11] Occurrences of negative κ values in other research have previously been reported.

of the coders agreeing completely in this respect); and two coders identified between three and five events.

As can be seen by the bar chart in figure D.1, the deviation from the mean κ value was slight — with the single exception of the above document. Strictly speaking, two cases of illegal grids were observed (that is, grids in which the event assignments did not rise sequentially). However, as we are only concerned with issues of event coreference, and not absolute position on the grid, this does not matter.

The graphical representations of the coders' segmentations gives us added confidence in the interpretation of the κ scores presented in section 9.6. In summary, we believe that the significant level of agreement among naive coders, and between the coders and the author, suggests that the segmentation approach we have adopted is intuitive and, therefore, justifiable as a discourse representation formalism.

9.12 Further issues

9.12.1 Commentary, Rhetoric and Politics

The integration of the CRaP document filter into the CONTESS framework made a significant improvement to the component scores, essentially limiting the degree to which the analysis modules react to spurious phrases that could otherwise be suggestive of multiple events.

Of course, as with all shallow approaches, there are bound to be exceptions where interesting (in the sense of genuinely eventful) documents are consequently excluded. The intersection of the set of eventful documents and the set of high CRaP-scoring documents is small. For example, one of the longer documents (txt78) in the test corpus is a transcription of a report from a rebel radio station in San Salvador. As such it contains many of the elements normally associated with the highly uneventful rhetoric produced by any politically motivated body, including frequent use of first and second person pronouns, terms like *comrades* and various salutations.

In retrospect, however, the CRaP filter was successful at identifying documents unlikely to contain multiple events, as the overall improved component scores show. One final modification that we would like to have made to the filter is the inclusion of punctuation — for example, the presence of an exclamation mark is an almost certain indication of a CRaP document (although again, not a guarantee that it does not contain multiple events).

9.12.2 Reported speech analysis

Although the approach to reported speech analysis implemented in the preprocessing component has not been formally evaluated, on inspection it appears to have performed well. For example, the following clauses would have been interpreted by the TAM as having the time stamp contained in the prefatory clauses.

- *fire department sources reported today that* guatemalan army troops killed nine peasants by mistake on 17 august.
- *in a statement released yesterday afternoon, the cuban ambassador regretted that* the united states is providing shelter for this kind of person and using them as "parrots to smear the cuba revolution."

However, there are also cases where important event distinction cues were missed due to RSA. In the first of the following examples, a temporal phrase was missed; in the second, a cue phrase.

- *castano was connected to the death of bernardo jaramillo, also a presidential candidate, for the leftist patriotic union (up), on 22 march, although at that time it was reported that ...*
- *meanwhile, medellin metropolitan police announced that*, early this morning...

The second of these examples was, however, identified as introducing a new event due to the presence of the temporal phrase.

Some modification of the range of reporting phrases used in RSA was undertaken during the development of CONTESS. For example, the inflected verb 'states' (as in 'the president states that...') was removed from the list of phrases due to the presence in the corpora of a large number of references to 'the united states'. Again, this is a problem that would be largely solved with a mixed-case corpus.

9.12.3 Corpus properties

If the results of the final evaluation are to be meaningful, then we must be able to assume that the test corpus is similar in both content and form to the development corpus. However, in order to ensure the unseen nature of the test corpus, we were unable to guarantee this prior to final evaluation. Of course, the assumption seems a fair one — the development and test corpus are taken from a previously used[12] larger corpus that was compiled specifically for the purposes of development followed by unseen evaluation.

The fact that this assumption was justified became clear during the manual segmentation and subsequent analysis of the test results. Nevertheless, very minor discrepancies between the two corpora can be found. For example, the numeric document analysis tables (sections C.4 and C.5) show that the average score obtained by proposing a minimally eventful hypothesis grid (denoted by [|]) for all documents in each corpus was 75.5 in the case of the development corpus, and 81.8 in the case of the test corpus. What this tells us is that the test corpus contains a lower proportion of eventful documents than the development corpus. This is confirmed by examining the average score obtained for the two corpora by proposing random grids[13] (denoted by [?]) which, as we saw in section 8.2.3, are generally quite eventful grids. In the case of the development corpus, the average score for random grids was 32.5; in the test corpus, 29.5.

9.13 Summary

This chapter has reported on the results of several system evaluations, including three evaluations of the development corpus (using the original grammar, using a modified grammar and using the CRaP document filter). Summary component analysis tables were presented that showed the contributions made by each aspect of the system towards correct segmentations. More detailed results tables, showing scores for individual documents and configurations respectively, were described (see appendix C). Results

[12] in the MUC-4 evaluation [DAR92]

[13] actually the average score obtained by generating 100 random grids

tables using tied ranked scores rather than numeric scores were also introduced.

The **results** section of the chapter presented the component analysis scores for each evaluation. A clear progression in scores was observed throughout the development evaluations, suggesting that rule evaluation and the adoption of document-level filtering are of benefit in automatic document segmentation. This progression is illustrated in figure 9.4, which compares the component scores across evaluations. The results of the naive coder agreement test were also described.

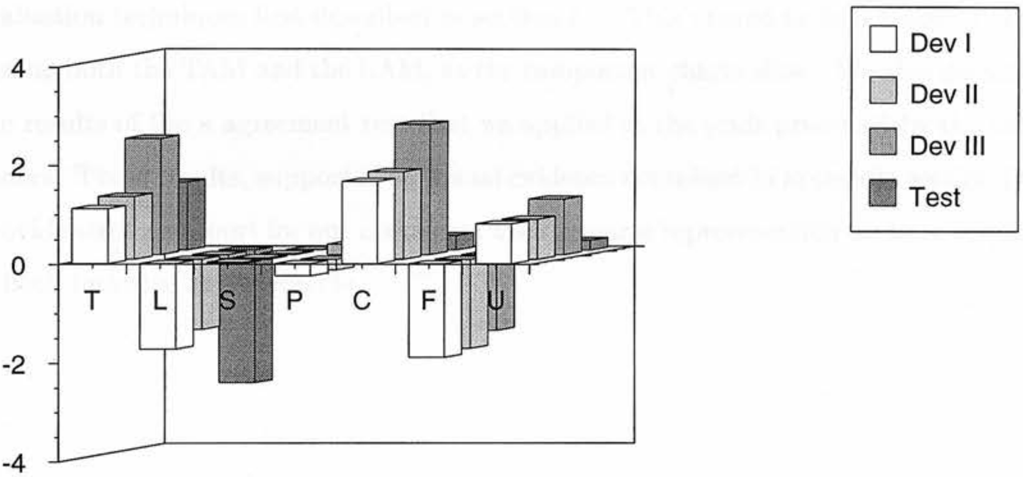


Figure 9.4: Comparison of component scores for all four evaluations

We then presented a **discussion** of the results, identifying where possible the reasons behind the observed performance of the analysis modules, gelling heuristics, rule evaluation, document filtering and miscellaneous other aspects of CONTESS. The TAM, we argued, is genuinely useful as a tool for distinguishing between multiple events in a text. The LAM, on the other hand, appears to be inappropriate for the task. Although rule evaluation and subsequent investigation showed that the rules in the LAM grammar, and the semantic interpretations that result from them, are satisfactory (in so far as locative phrases are readily identified and interpreted), it simply seems that locative phrases are not useful indicators of event shifts in the same way that temporal phrases are. The CPAM was continually positive in its contribution to event distinction, contrasting with Iwańska's findings [Iwańska 93] that cue phrases are unreliable indicators of event structure.[14] In fact, as can be seen in figure 9.4, the component

[14] At least part of this discrepancy stems from varying definitions of what constitutes a cue phrase.

score profile of the CPAM closely mirrors that of the TAM in terms of proportional scores across evaluations. Due to their dependence on the analysis modules to propose correct event shifts, the performance of the sentence and, in particular, the clause gelling heuristics varied accordingly. Use of the term vector frequency analysis component by the Event Manager adversely affected performance in all but the final test corpus analysis, a result that we are unable to fully explain.

Finally, a number of other features of the evaluation were discussed, including the rule evaluation techniques first described in section 8.4. This proved to be a useful tool for tuning both the TAM and the LAM, as the component charts show. We also discussed the results of the κ agreement test that we applied to the grids produced by the naive coders. These results, supported by visual evidence contained in appendix section D.3, provide strong support for our claim that the discourse representation we have adopted is both intuitive and workable.

Chapter 10

Conclusions

At the beginning of this thesis we outlined several aims. Foremost among these was the investigation of the potential discourse structuring information content of a range of linguistic phenomena, and the feasibility of using shallow processing techniques based on these phenomena for discourse segmentation purposes. In order to carry out these investigations, we required a representation formalism, an evaluation methodology and a modular framework to test our theories. All these aims have now been met. In this concluding chapter, we shall present the contributions that have been made by the thesis, discuss some of the issues that it has raised, and suggest some areas for further work.

10.1 Contributions of the thesis

The first contribution of this thesis is the clause-event grid **formalism** for representing multi-event texts. The main strengths of the formalism are that it is descriptive, in that it focuses on relationships between events rather than requiring the specification of an absolute position; it is plausible, as has been shown by our experiments with naive human coders; and it is able to be evaluated.

The **methodology** for quantitatively evaluating the grid representations constitutes the second of the contributions made by the thesis. We have argued that dichotomic metrics, such as recall and precision, are not well suited to the problem of evaluating relational discourse segmentations, and have proposed a methodology for comparing clause-event grids that is both intuitive and productive.

Having established a technique for evaluating grid representations, our third contribution is the **modular framework** of CONTESS, which allows the analysis modules and event manager, with its various heuristics, to be used as a tool for investigating the potential benefits of processing a range of linguistic phenomena for text structuring. The fourth contribution is therefore the **insight** into the possible value of using specific linguistic phenomena for this purpose and, in particular, the conclusion that temporal information (as extracted by the TAM) is consistently useful for distinguishing between multiple events in a text (as opposed to locative information, which appears to be inappropriate for the task). The success of the TAM (and, to a lesser degree, the CPAM and the clause gelling heuristic) suggests that shallow NLP approaches represent a realistic solution to text segmentation problems.

10.2 Issues raised by the thesis

A number of further issues have been raised in the course of this thesis. Among these are the use of statistical rule evaluation techniques, domain dependence, and the usability of clause-event grids.

10.2.1 Rule evaluation

During the development of the grammars for the temporal and locative analysis modules, we became interested in the prospect of automatically scoring rules as an *aid* to grammar improvement. This proved useful in two ways. Firstly, it helped us to identify rules that were adversely affecting the performance of their respective module. Secondly, it led to the realisation that, for certain rules, frequency of application and success of application (in so far as the interpretations were as we had intended) were not necessarily evidence of those rules' utility for event recognition. This was particularly true in the case of rules in the LAM.

10.2.2 Domain dependence

Throughout the course of our investigation into event recognition mechanisms, we have been conscious of the ease with which such mechanisms can become highly domain-dependent. For example, the simplification of the concept of an event to one of a fixed set of incident types restricts any ER approach to being domain-dependent. Whilst we admit to targeting our investigation towards a particular genre of domains, i.e. physically oriented instantaneous events, we have been careful to avoid designing any of the analysis modules specifically for Latin American terrorist incidents.[1] Furthermore, we believe that the genre of domains that we are concerned with is sufficiently wide to make our findings of general interest.

This is not to say that there is anything wrong *per se* with domain-dependent ER techniques. Indeed, as has been the case in the broader field of IE, domain-dependent approaches have often yielded the best results, with automatic pattern acquisition techniques (as seen in Riloff's aptly named AutoSlog [Riloff 93] and Soderland's subsequent Crystal [Soderland *et al.* 95]) meaning that, with the availability of tagged training data, the construction of domain-specific rules is no longer a prohibitively labour-intensive task. Although a comparison of domain-dependent and -independent approaches would no doubt be of interest, it remains outside the scope of this thesis. We would, however, consider the inclusion of a domain-dependent topic spotting module as a possible extension, as we explain below in section 10.3.

10.2.3 Usability of clause-event grids

The usability of our discourse structure representation formalism is important if our results are to be meaningful. If it were the case that human subjects were unable to map their perceived discourse structure to a clause-event grid, then the usability of the representation would be called into question. This would also suggest that either the status of clauses or events (as the units of the grid axes) was unclear, or that the relationship between them on the grid was not workable.

[1] The restriction of the LAM knowledge base and keywords to only Latin American place names was done purely for reasons of scale.

As the results of our human coder agreement experiments show, the usability of the clause-event grids is not an area for concern. However, this is not to say that the grids are a completely satisfactory means of representing the distribution of events in a text. Under the current definition of a clause-event grid, the relationship between clauses and events is one-to-one. That is, it is not possible to show single clauses identifying with multiple events. This is the inevitable trade-off between descriptive potential and evaluability. Having said this, we do not feel that the loss of descriptive potential is serious. Our experience with the corpus suggests that the occasions when the event structure of a text extends beyond the singular relationship available in clause-event grids are relatively few. This is backed up by Iwańska's observation that SI (shared information) structures are relatively rare (see section 2.3).

10.3 Further work

In section 3.6.1 we listed six elements that have been often cited as useful for performing event distinction. Three of them (time, location and cue phrases) were implemented within CONTESS. Of the rest (topic, tense and aspect), the identification of topic, i.e. incident type in a MUC context, is the most obvious extension. Besides allowing a more complete investigation of discourse segmentation cues, the endowment of a "topic spotting" ability to CONTESS would appear to have two significant benefits. Firstly, the concept of an event is clarified enormously, as an event becomes one of a fixed set of incident types (e.g. *arson*). Secondly, there then follows the tantalising prospect of performing a more rigorous and extensive evaluation using the MUC templates, which are available for each document in the MUC development corpus on the basis of one template per relevant event (i.e. incident type).

However, the integration of a topic spotting ability into CONTESS was not adopted for a number of reasons. Although the reduction of an event to one of a set of incident types is attractive for the sake of simplification, we feel that the adoption of a strictly domain-dependent definition of an event imposes excessive restrictions on the wider applicability of our approach. Furthermore, the feasibilty of using the MUC templates for extensive evaluation purposes is not clear. If we are to attempt to evaluate anything more than a program's judgement of how *many* events a document contains, it becomes necessary to relate MUC templates to specific fragments of the text. This is difficult to do with any accuracy, although it has been achieved on an approximate basis, as with Cowie *et al.*'s topic spotting component [Cowie *et al.* 92]. Nevertheless, the implementation and evaluation of a topic spotting program as a module alongside CONTESS's currently existing ones remains an obvious extension.

Although certain changes in tense and aspect are claimed to act as event shift cues [Iwańska *et al.* 91], there is little evidence in the literature of their use for discourse segmentation purposes. In cases where a series of events is being described in the past tense (e.g. in our example text), it is not clear that tense changes are necessarily reliable indications of event shifts. For example, the following paragraph describes one event; however, there is a change of tense in the matrix clause of the second sentence from the past to the present tense.

> A bomb exploded yesterday in downtown Aracataca. Police in Bogota say that the JPF were responsible for the attack. Several buildings were damaged in the blast.

On the other hand, it may be that reported speech analysis is useful in this sort of situation. Changes in the tense of non-reporting verbs, such as *exploded*, e.g. to *had exploded*, might represent a more reliable indication. In order to accurately identify verbs, at least some degree of part of speech tagging would be required. Given the speed, accuracy and free availability of such taggers, however, this is unlikely to be a problem. Investigation of the relationship between tense and aspect changes and event shifts is therefore one area that we would be interested in pursuing.

On a more practical note, the parallelisation of CONTESS's analysis modules is a desirable extension — especially if further modules (such as topic spotting, tense and aspect) are to be considered. There is no theoretical reason why this cannot be achieved, which is why we have presented the analysis modules as parallel processes in diagrams showing the architecture of CONTESS. As a result of parallelisation, overall runtime would then be reduced from the sum of the runtimes of individual modules to the runtime of the single slowest module.

10.4 Summary

Although it is unlikely that Goethe had information extraction in mind when he delivered the quote at the start of this thesis, IE systems would appear to offer one of the most potent solutions to the growing information overload problem. However, unless IE researchers confront the pressing discourse-related issues that continue to restrict progress in the field, IE will remain fundamentally unable to deal with structurally complex documents.

In this thesis we have proposed a mechanism for evaluating approaches to event recognition, including a formalism for representing event structure and a constraint based platform for testing such approaches. Using the resulting CONTESS system, we have shown that certain linguistic phenomena, such as temporal phrases and cue phrases, are useful indicators of event structure, whereas others, such as locative phrases, are not. Although there are other linguistic phenomena and structuring heuristics still to evaluate, our findings suggest that shallow event recognition techniques constitute a promising approach to discourse segmentation, and the quantitative methodology that we have proposed facilitates further research in the field.

Bibliography

[Andrews & Peterson 90] P P Andrews and M B Peterson, editors. *Criminal Intelligence Analysis*. Palmer Enterprises, Loomis, CA, 1990.

[Aone *et al.* 92] Chinatsu Aone, Doug McKee, Sandy Shinn, and Hatte Blejer. Description of the Solomon system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 259—267. DARPA, Morgan Kaufmann, 1992.

[ARP94] ARPA. *Proceedings of the fifth Message Understanding Conference, MUC-5*, San Mateo, CA, 1994. Morgan Kaufmann.

[ARP96] ARPA. *Proceedings of the sixth Message Understanding Conference, MUC-6*, San Francisco, CA, April 1996. Morgan Kaufmann.

[August & Dolan 92a] Stephanie E August and Charles P Dolan. Description of the trainable text skimmer used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 189—196. DARPA, Morgan Kaufmann, 1992.

[August & Dolan 92b] Stephanie E August and Charles P Dolan. Hughes trainable text skimmer: MUC-4 test results and analysis. In *Proceedings of the fourth Message Understanding Conference, MUC-4*. DARPA, Morgan Kaufmann, 1992.

[Ayuso 93] Damaris M Ayuso. Topic session on discourse. In *Proceedings of the fifth Message Understanding Conference, MUC-5*, page 345, San Francisco, CA, August 1993. ARPA, Morgan Kaufmann.

[Ballard *et al.* 71] D Ballard, R Conrad, and R Longacre. The deep and surface grammar of interclausal relations. *Foundations of Language*, 4:70—118, 1971.

[Belvin *et al.* 89] R S Belvin, H K Holmback, C A Montgomery, B G Stalls, and R E Stumberger. Message fusion: final report. Technical Report OSD-W-R88-06, Logicon Inc., Operating Systems Division, 1989.

[Carletta 95] Jean Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 30(11), 1995.

[Chapman *et al.* 70] Graham Chapman, John Cleese, Terry Gilliam, Eric Idle, Terry Jones, and Michael Palin. Monty Python's Flying Circus, episode 20, The Attilla The Hun Show. BBC TV, November 1970.

[Chinchor 91] Nancy Chinchor. MUC-3 evaluation metrics. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 17—24, San Mateo, CA, May 1991. DARPA, Morgan Kaufmann.

[Chinchor *et al.* 93] Nancy Chinchor, Lynette Hirschman, and David D Lewis. Evaluating message understanding systems: an analysis of the Third Message Understanding Conference (MUC-3). *Computational Linguistics*, 19(3):409—449, September 1993.

[Cleverdon 91] Cyril W Cleverdon. The significance of Cranfield tests of indexed languages. In *Proceedings of the 14th International Conference on Research and Development in Information Retrieval*, 1991.

[Coates-Stephens 92] Sam Coates-Stephens. *The analysis and acquisition of proper names for robust text understanding.* Unpublished PhD thesis, City University, October 1992.

[Cohen 84] R Cohen. A computational theory of the function of clue words in argument understanding. In *COLING-84*, pages 251—258, 1984.

[Cowie *et al.* 92] Jim Cowie, Louise Guthrie, Yorick Wilks, James Pustejovsky, and Scott Waterman. Description of the MucBruce system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 223—232. DARPA, Morgan Kaufmann, 1992.

[Creary *et al.* 89] Lewis G Creary, J Mark Gawron, and John Nerbonne. Reference to locations. In *Proceedings of the 27th annual meeting of the Association for Computational Linguistics*, pages 42—50. Association for Computational Linguistics, June 1989.

[Crowe 93] Jeremy Crowe. Event recognition in information extraction. Discussion paper 137, Department of Artificial Intelligence, Edinburgh University, Edinburgh, UK, August 1993.

[Cullingford 78] Richard E Cullingford. *Script application: computer understanding of newspaper stories.* Unpublished PhD thesis, Yale University, New Haven, CT, 1978.

[Cullingford 86] Richard E Cullingford. Sam. In Barbara J Grosz, Karen Sparck Jones, and Bonnie Lynn Webber, editors, *Readings in natural language processing*, pages 627—649. Morgan Kaufmann, 1986.

[Cunningham *et al.* 95] Hamish Cunningham, Robert J Gaizauskas, and Yorick Wilks. A General Architecture for Text Engineering (GATE): A New Approach to Language Engineering R&D. Technical report CS-95-21, Department of Computer Science, Sheffield University, 1995.

[Dahlgren *et al.* 91] Kathleen Dahlgren, Carol Lord, Hajime Wada, Joyce McDowell, and Edward P Stabler Jr. Description of the Interpretext system as used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 163—170. DARPA, Morgan Kaufmann, 1991.

[DAR91] DARPA. *Proceedings of the third Message Understanding Conference, MUC-3*, San Mateo, CA, 1991. Morgan Kaufmann.

[DAR92] DARPA. *Proceedings of the fourth Message Understanding Conference, MUC-4*, San Mateo, CA, 1992. Morgan Kaufmann.

[DeJong 79] Gerald DeJong. *Skimming stories in real time.* Unpublished PhD thesis, Yale University, New Haven, CT, 1979.

[Dietz 91] Robert Dietz. Description of the TIA system used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 150—154. DARPA, Morgan Kaufmann, 1991.

[Dolan *et al.* 91] Charles P Dolan, Thomas V Cuda, Seth R Goldman, and Alan M Nakamura. Description of the trainable text skimmer used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 155—162. DARPA, Morgan Kaufmann, 1991.

[Evans] Roger Evans. Personal communication.

[Fleischman 90] S Fleischman. *Tense and narrativity.* University of Texas Press, Austin, TX, 1990.

[GECS 92] GECS. NLToolset: More Than Just Words. General Electric Consulting Services product release, 1992.

[Gould 71] H W Gould. Research bibliography of two special number sequences. *Mathematica Monongoliae*, 12, 1971.

[Grishman *et al.* 92] Ralph Grishman, Catherine Macleod, and John Sterling. Description of the PROTEUS system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 233—241. DARPA, Morgan Kaufmann, 1992.

[Grosz & Sidner 86] Barbara Grosz and Candy Sidner. Attention, intensions and the structure of discourse. *Computational Linguistics*, 12(3), 1986.

[Halliday & Hasan 76] M Halliday and R Hasan. *Cohesion in English.* Longman, Harlow, Essex, 1976.

[Hearst 94] Marti A Hearst. Multi-paragraph segmentation of expository text. In *Proceedings of the 32nd meeting of the Association for Computational Linguistics*, Las Cruces, NM, 1994. ACL.

[Hirschberg & Litman 93] Julia Hirschberg and Diane Litman. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3), 1993.

[Hirschman & Story 81] Lynette Hirschman and Guy Story. Representing implicit and explicit time relations in narrative. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 289—295, 1981.

[Hirschman 92] Lynette Hirschman. An adjunct test for discourse processing in MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 67—77. DARPA, Morgan Kaufmann, 1992.

[Hirst 81] Graeme Hirst. *Anaphora in natural language understanding: a survey.* Lecture notes in computer science. Springer-Verlag, 1981.

[Hobbs 86] Jerry R Hobbs. Resolving pronoun references. In Barbara J Grosz, Karen Sparck Jones, and Bonnie Lynn Webber, editors, *Readings in natural language processing*, pages 339—352. Morgan Kaufmann, Los Altos, 1986. Article first published in 1978.

[Hobbs 91] Jerry R Hobbs. SRI International's TACITUS system: MUC-3 test results and analysis. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 105—107. DARPA, Morgan Kaufmann, 1991.

[Hobbs *et al.* 90] Jerry R Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. Interpretation as abduction. Technical Report 499, SRI, 1990.

[Hobbs *et al.* 91] Jerry R Hobbs, Douglas E Appelt, John S Bear, Mabry Tyson, and David Magerman. The TACITUS system. Technical Report 511, SRI, 1991.

[Hobbs *et al.* 92a] Jerry R Hobbs, Douglas E Appelt, John S Bear, David J Israel Sr, and Mabry W Tyson. FASTUS: A system for extracting information from natural-language text. Technical Report 519, SRI, 1992.

[Hobbs *et al.* 92b] Jerry R Hobbs, Douglas E Appelt, John S Bear, David J Israel Sr, and Mabry W Tyson. Description of the FASTUS system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 268—275. DARPA, Morgan Kaufmann, 1992.

[Iwańska 91] Lucja Iwańska. Discourse processing module. Technical Information Series 91CRD180, Artificial Intelligence Laboratory, GE Research and Development Center, September 1991.

[Iwańska 93] Lucja Iwańska. Discourse structure in factual reports. Unpublished, 1993.

[Iwańska *et al.* 91] Lucja Iwańska, Douglas Appelt, Damaris Ayuso, Kathy Dahlgren, Bonnie Glover Stalls, Ralph Grishman, George Krupka, Christine Montgomery, and Ellen Riloff. Computational aspects of discourse in the context of MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 256—282. DARPA, Morgan Kaufmann, 1991.

[Jacobs *et al.* 91] Paul S Jacobs, George R Krupka, and Lisa F Rau. Lexico-semantic pattern matching as a companion to parsing in text understanding. In *Fourth DARPA Speech and Natural Language Workshop*, San Mateo, CA, February 1991. DARPA, Morgan Kaufmann.

[Kamp 81] Hans Kamp. A theory of truth and semantic interpretation. In J Gronendijk, T Janssen, and M Stokhof, editors, *Formal methods in the study of language*. Mathematisch Centrum, Amsterdam, 1981.

[Knott 95] Alistair Knott. *A Data-Driven Methodology for Motivating a Set of Coherence Relations.* Unpublished PhD thesis, Edinburgh University, Edinburgh, UK, 1995.

[Kozima & Furugori 93] Hideki Kozima and Teiji Furugori. Similarity between words computed by spreading activation on an English dictionary. In *Proceedings of the 1993 meeting of the European Association for Computational Linguistics*, 1993.

[Kozima 93] Hideki Kozima. Text segmentation based on similarity between words. In *Proceedings of the 31st meeting of the Association for Computational Linguistics, student session.* ACL, 1993.

[Krippendorff 80] Klaus Krippendorff. *Content analysis: an introduction to its methodology.* Sage Publications, 1980.

[Krupka *et al.* 91] George R Krupka, Paul S Jacobs, Lisa F Rau, and Lucja Iwańska. Description of the NLToolset system as used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 144—149. DARPA, Morgan Kaufmann, 1991.

[Krupka *et al.* 92] George Krupka, Paul Jacobs, and Lisa Rau. GE NLToolset: Description of the system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 177—185. DARPA, Morgan Kaufmann, 1992.

[Kuhns 88] Robert J Kuhns. A news analysis system. In *Computational Linguistics*, 1988.

[Labov 72] William Labov. *Language in the inner city : studies in the Black English vernacular*, chapter The transformation of experience in narrative syntax. University of Pennsylvania Press, 1972.

[LDOCE 87] *Longman Dictionary of Contemporary English.* Harlow, Essex, new edition, 1987.

[Lebowitz 80] Michael Lebowitz. *Generalization and memory in an integrated understanding system.* Unpublished PhD thesis, Yale University, New Haven, CT, 1980.

[Lehnert 94] Wendy G Lehnert. Cognition, Computers, and Car Bombs: How Yale Prepared Me for the 90's. In Schank and Langer, editors, *Beliefs, Reasoning and Decision Making: Psychologic in Honor of Bob Abelson*, pages 143—173. Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.

[Lehnert *et al.* 91] Wendy Lehnert, Claire Cardie, David Fisher, Ellen Riloff, and Robert Williams. Description of the CIRCUS system as used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 223—233. DARPA, Morgan Kaufmann, 1991.

[Lehnert *et al.* 92] W Lehnert, C Cardie, D Fisher, J McCarthy, E Riloff, and S Soderland. Description of the CIRCUS system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 282—288. DARPA, Morgan Kaufmann, 1992.

[Lehnert *et al.* 95] W Lehnert, C Cardie, D Fisher, J McCarthy, E Riloff, and S Soderland. Evaluating an information extraction system. In *Journal of integrated computer-aided engineering*, pages 453—472. Wiley, 1995.

[Lewis & Tong 92] David D Lewis and Richard M Tong. Text filtering in MUC-3 and MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 51—66. DARPA, Morgan Kaufmann, 1992.

[Loatman 92] Bruce Loatman. Description of the PAKTUS system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 253—258. DARPA, Morgan Kaufmann, 1992.

[Lytinen *et al.* 92] Steven L Lytinen, Sayan Bhattacharyya, Robert R Burridge, Peter M Hastings, Christian Huyck, Karen A Lipinsky, Eric S McDaniel, and Karenann K Terrel. Description of the link system used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 289—295. DARPA, Morgan Kaufmann, 1992.

[Mann & Thompson 88] W C Mann and S A Thompson. Rhetorical structure theory: A theory of text organisation. *Text*, 8(3):243–281, 1988. Also available as Tech Report RR-87-190, USC Information Sciences Institute, Marina del Rey, CA.

[McDonald 93] D McDonald. Internal and external evidence in the identification and semantic categorization of proper names. In *Proceedings of ACL SIGLEX Workshop on Lexical Acquisition from Text*, 1993. Reprinted in *Corpus Processing for Lexical Acquisition*, B Boguraev and J Pustejovsky (eds), MIT Press, 1996.

[Meyers & deHilster 92] Amnon Meyers and David de Hilster. Description of the TEXUS system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 207—214. DARPA, Morgan Kaufmann, 1992.

[Moens 92] Marc Moens. Temporal order and discourse type. In Steffen Leo Hansen and Finn Sørensen, editors, *Topics in Sematic Interpretation*, pages 55—84. Sammfundsliteratur, 1992.

[Montgomery & Glover 84] C A Montgomery and B C Glover. A text grammar for automated analysis of C3I messages. In *Proceedings of the 1984 Conference on Intelligent Systems and Machines*, pages 347—352, Oakland University, Rochester, MI, 1984.

[Montgomery *et al.* 89] C A Montgomery, J Burge, H Holmback, J L Kuhns, B G Stalls, R Stumberger, and R L Russel Jr. The DBG Message Understanding System. In *Proceedings of the Annual*

AI Systems in Government conference, pages 258—265, Washington, DC, 1989. Computer Society of the IEEE.

[Montgomery *et al.* 91] Christine A Montgomery, Bonnie Glover Stalls, Robert S Belvin, and Robert E Stumberger. Language Systems Inc: Description of the DBG system as used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 171—177. DARPA, Morgan Kaufmann, 1991.

[Montgomery *et al.* 92] Christine A Montgomery, Bonnie Glover Stalls, Robert E Stumberger, Naicong Li, Robert S Belvin, Alfredo Arnaiz, and Susan B Hirsh. Language Systems Inc: Description of the DBG system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 197—206. DARPA, Morgan Kaufmann, 1992.

[Montgomery *et al.* 93] Christine A Montgomery, Robert E Stumberger, Bonnie Glover Stalls, Naicong Li, Robert S Belvin, and Susan Hirsh Litenatsky. Description of the DBG system as used for MUC-5. In *Proceedings of the fifth Message Understanding Conference, MUC-5*, pages 121—135, San Francisco, August 1993. ARPA, Morgan Kaufmann.

[Morris & Hirst 91] Jane Morris and Graeme Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1):21—42, 1991.

[Moser & Moore 95] Megan Moser and Johanna Moore. Using discourse analysis and automatic text generation to study discourse cue usage. In *Proceedings of the AAAI spring symposium on empirical methods in discourse interpretation and generation*, pages 92—98, Stanford, CA, 1995.

[Noel 86] Dirk Noel. *Towards a functional characterization of the news of the BBC World Service*, volume 49 of *Antwerp papers in linguistics*. Wilrijk, University of Antwerp, 1986.

[NRaD 92] NRaD. MUC-4 task documentation. NRaD (Naval Research and Development), NCCOSC On-line document, 1992.

[Nunberg 90] Geoffrey Nunberg. *The linguistics of punctuation*. CSLI lecture notes; no. 18. Center for the study of language and information, 1990.

[Paice 90] Chris Paice. Constructing literature abstracts by computer: Techniques and prospects. *Journal of Information Processing and Management*, 26(1):171—186, 1990.

[Porter 80] M F Porter. An algorithm for suffix stripping. *Program*, 14(3):130—137, 1980.

[Propp 68] Vladimir Iakovlevich Propp. *Morphology of the folktale.* University of Texas Press, Austin, TX, 1968. Translated by L Scott. Findings originally published in 1928.

[Quirk *et al.* 72] R Quirk, S Greenbaum, G Leech, and J Svartik. *A grammar of contemporary English.* Longmans, London, 1972.

[Rapaport *et al.* 89] William Rapaport, Erwin Segal, Stuart Shapiro, David Zubin, Gail Bruder, Judith Duchan, and David Mark. Cognitive and computer systems for understanding narrative texts. Technical report 89-07, Department of Computer Science, SUNY, Buffalo, NY, 1989.

[Rau & Jacobs 88] Lisa F Rau and Paul S Jacobs. Integrating top-down and bottom-up strategies in a text processing system. In *Proceedings of the second conference on applied natural language processing*, pages 129—135, Morristown, NJ, Feb 1988. ACL.

[Reichman 85] R Reichman. *Getting computers to talk like you and me: discourse context, focus, and semantics.* The MIT Press, 1985.

[Rigler 92] Elina Rigler. *The spatio-temporal structure of narrative texts: a study of aspect.* Unpublished PhD thesis, Edinburgh University, Edinburgh, UK, 1992.

[Riloff 93] Ellen Riloff. Automatically constructing a dictionary for information extraction tasks. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 811—816, Washington, DC, July 1993. MIT Press.

[Sager 78] Naomi Sager. Natural language information formatting: The automatic conversion of texts to a structured database. In M C Yovits, editor, *Advances in Computers.* Academic Press, New York, 1978.

[Sager 81] Naomi Sager. *Natural Language Information Processing: A computer grammar of English and its applications*, chapter 8, pages 213—231. Addison-Wesley, 1981.

[Salton 71] Gerald Salton, editor. *The Smart Information Retrieval System — Experiments in Automatic Document Processing.* Prentice Hall, Inc, Englewood Cliffs, NJ, 1971.

[Schank & Abelson 77] Roger C Schank and Robert P Abelson. *Scripts, plans, goals and understanding.* Erlbaum, Hillsdale, NJ, 1977.

[Schank 72] Roger C Schank. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 3:552—631, 1972.

[Schank 80] Roger C Schank. Language and memory. *Cognitive Science*, 4(3):243—284, 1980.

[Schank 86] Roger C Schank. Intelligent advisory systems. In Patrick H Winston and Karen A Prendergast, editors, *The AI Business: commercial applications of artificial intelligence*, chapter 11, pages 133—146. MIT, 1986.

[Schiffrin 87] Deborah Schiffrin. *Discourse markers.* Studies in interactional sociolinguistics. Cambridge University Press, Cambridge, UK, 1987.

[Schourup 85] Lawrence C Schourup. *Common discourse particles in English conversation.* Outstanding dissertations in linguistics. Garland Publishing, 1985.

[Senator *et al.* 95] Ted E Senator, Henry G Goldberg, Jerry Wooton, Matthew A Cottini, A F Umar Khan, Christina D Klinger, Winston M Llamas, Michael P Marrone, and Raphael W H Wong. The Financial Crimes Enforcement Network AI System (FAIS). *AI Magazine*, pages 21—39, Winter 1995.

[Shieber *et al.* 83] S M Shieber, H Uszkoreit, F C N Pereira, J Robinson, and M Tyson. The formalism and implementation of patr-ii. in *research on interactive acquisition and use of knowledge.* volume 1894 of sri final report. Technical report, SRI International, Menlo Park, 1983.

[Sidner 81] Candace L Sidner. Focusing for interpretation of pronouns. *American Journal of Computational Linguistics*, 7(4):217—231, 1981.

[Siegel & Castellan 88] Sidney Siegel and N John Castellan. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, London, 2nd edition, 1988.

[Sloane & Plouffe 95] N J A Sloane and S Plouffe. *The Encyclopedia of Integer Sequences.* Academic Press, San Diego, 1995.

[Sloane 94] N J A Sloane. The online encyclopedia of integer sequences. *The Electronic Journal of Combinatorics*, 1, 1994. Online at `http://ejc.math.gatech.edu:8080/Journal/`.

[Soderland *et al.* 95] Stephen Soderland, David Fisher, Jonathan Aseltine, and Wendy Lehnert. Issues in inductive learning of domain-specific text extraction rules. In *Proceedings of the IJCAI-95 workshop on new approaches to learning for natural language processing*, 1995.

[Stalls *et al.* 90] B G Stalls, R Stumberger, and C A Montgomery. Long Range Air (LRA) Data Base Generator (DBG) final technical report. Technical Report RADC-TR-89-366, Rome

Air Development Center, Grifiss Air Force Base, Rome, NY, 1990.

[Sundheim 91] Beth M Sundheim. Overview of the third message understanding conference. In *Proceedings of the Third Message Understanding Conference, MUC-3*, pages 3—16. DARPA, Morgan Kaufmann, 1991.

[Sundheim 92] Beth M Sundheim. Overview of the fourth message understanding conference. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 3—21. DARPA, Morgan Kaufmann, 1992.

[Sundheim 95] Beth Sundheim. Named entity task definition. Online document — contact author for current URL, May 1995. http://cs.nyu.edu/cs/faculty/grishman/muc6.html.

[Tarantino 92] Quentin Tarantino. *Reservoir Dogs.* Miramax Films, 1992.

[van Dijk 72] T A van Dijk. *Some Aspects of Text Grammar.* Mouton, The Hague, 1972.

[Van Rijsbergen 79] C J Van Rijsbergen. *Information retrieval.* Butterworths, London, 2nd edition, 1979.

[vanDijk & Kintsch 83] T A van Dijk and W Kintsch. *Strategies of Discourse Comprehension.* Academic Press, 1983.

[Vilain *et al.* 96] Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. A model-theoretic coreference scoring scheme. In *Proceedings of the sixth Message Understanding Conference, MUC-6*, pages 45—52, San Francisco, CA, 1996. ARPA, Morgan Kaufmann.

[Wakao *et al.* 96] T Wakao, R Gaizauskas, and Y Wilks. Evaluation of an algorithm for the recognition and classification of proper names. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, pages 418—423, Copenhagen, 1996.

[Wall & Schwartz 91] Larry Wall and Randal L Schwartz. *Programming Perl.* O'Reilly and Associates, Inc, 1991.

[Weir & Silk 92] Carl Weir and Barry Silk. Description of the Paramax system as used for MUC-4. In *Proceedings of the fourth Message Understanding Conference, MUC-4*, pages 242—252. DARPA, Morgan Kaufmann, 1992.

[Weir *et al.* 91] Carl Weir, Tim Finin, Robin McEntire, and Barry Silk. Description of the Unisys system used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 212—222. DARPA, Morgan Kaufmann, 1991.

[Weischedel *et al.* 91] Ralph Weischedel, Damaris Ayuso, Sean Boisen, Robert Ingria, and Jeff Palmucci. Description of the PLUM system as used for MUC-3. In *Proceedings of the third Message Understanding Conference, MUC-3*, pages 137—143. DARPA, Morgan Kaufmann, 1991.

[Wilensky 78] Robert Wilensky. *Understanding goal-based stories.* Unpublished PhD thesis, Yale University, New Haven, CT, 1978.

[Young & Hayes 85] Sheryl R Young and Philip J Hayes. Automatic classification and summarization of banking telexes. In *Proceedings of the Second Conference on Artificial Intelligence Applications: the engineering of knowledge-based systems*, pages 402—408, Miami Beach, FL, December 1985. IEEE Computer Society.

[Zuckerman & Pearl 86] Ingrid Zuckerman and Judea Pearl. Comprehension-driven generation of meta-technical utterances in math tutoring. In *Proceedings of the fifth national conference of the AAAI*, pages 606—611, Philadelphia, PA, 1986.

Appendix A

Publications

- Shallow techniques for the segmentation of news reports. In *Proceedings of the 1996 AISB workshop on language engineering for document analysis and recognition*, AISB, April 1996.
- Constraint based event recognition for information extraction. In *Proceedings of the 33rd meeting of the Association for Computational Linguistics, student session*, ACL, June 1995.
- *Event recognition in information extraction*, Department of Artificial Intelligence discussion paper no. 137, Edinburgh University, August 1993.
- *Multicomponent text processing*, unpublished MPhil thesis, Cambridge University, August 1992.

Appendix B

Detailed example

This appendix contains the output produced by each component of CONTESS when presented with an example document. The document used is from our development corpus, and has also been used in the MUC-3 evaluation [DAR91] as a demonstration text.

We begin by showing the original document as encountered by CONTESS, and then present the output of the preprocessor. The next three sections contain the phrases identified, their interpreted forms, and the constraints produced by each analysis module. The final section shows the grids output by the event manager in each of its 128 configurations, and illustrates the constraint relationships that exist in the document using a link-analysis visualisation technique.

It should be noted that the version of CONTESS used in this example is the same as used with the test corpus — i.e. modified TAM grammar and CRaP filter selected (although this document is assigned a low CRaP score).

B.1 Original text

```
TST1-MUC3-0099

   LIMA, 25 OCT 89 (EFE) -- [TEXT] POLICE HAVE REPORTED THAT
TERRORISTS TONIGHT BOMBED THE EMBASSIES OF THE PRC AND THE SOVIET
UNION.  THE BOMBS CAUSED DAMAGE BUT NO INJURIES.

   A CAR-BOMB EXPLODED IN FRONT OF THE PRC EMBASSY, WHICH IS IN THE
LIMA RESIDENTIAL DISTRICT OF SAN ISIDRO.  MEANWHILE, TWO BOMBS WERE
THROWN AT A USSR EMBASSY VEHICLE THAT WAS PARKED IN FRONT OF THE
```

```
EMBASSY LOCATED IN ORRANTIA DISTRICT, NEAR SAN ISIDRO.

   POLICE SAID THE ATTACKS WERE CARRIED OUT ALMOST SIMULTANEOUSLY AND
THAT THE BOMBS BROKE WINDOWS AND DESTROYED THE TWO VEHICLES.

   NO ONE HAS CLAIMED RESPONSIBILITY FOR THE ATTACKS SO FAR. POLICE
SOURCES, HOWEVER, HAVE SAID THE ATTACKS COULD HAVE BEEN CARRIED OUT BY
THE MAOIST "SHINING PATH" GROUP OR THE GUEVARIST "TUPAC AMARU
REVOLUTIONARY MOVEMENT" (MRTA) GROUP.  THE SOURCES ALSO SAID THAT THE
SHINING PATH HAS ATTACKED SOVIET INTERESTS IN PERU IN THE PAST.

   IN JULY 1989 THE SHINING PATH BOMBED A BUS CARRYING NEARLY 50
SOVIET MARINES INTO THE PORT OF EL CALLAO. FIFTEEN SOVIET MARINES WERE
WOUNDED.

   SOME 3 YEARS AGO TWO MARINES DIED FOLLOWING A SHINING PATH BOMBING
OF A MARKET USED BY SOVIET MARINES.

   IN ANOTHER INCIDENT 3 YEARS AGO, A SHINING PATH MILITANT WAS KILLED
BY SOVIET EMBASSY GUARDS INSIDE THE EMBASSY COMPOUND.  THE TERRORIST
WAS CARRYING DYNAMITE.

   THE ATTACKS TODAY COME AFTER SHINING PATH ATTACKS DURING WHICH
LEAST 10 BUSES WERE BURNED THROUGHOUT LIMA ON 24 OCT.
```

As we can see, this document is a highly eventful one containing, in our view, 6 distinct events:

- car bomb attack against the PRC embassy
- bombing of the Soviet embassy
- attack against a bus in July 1989
- attack on a market, killing two marines
- terrorist killed by Soviet embassy guards
- arson attack on buses in Lima

Having said that SI structures [Iwańska 93] (where two events are described by the same clause) are rare, it just so happens that this document does indeed contain an instance. The attacks against the two embassies described at the start of the text are described together. In this case, the solution is to interpret this as one event, until the point that the two events are differentiated above the clause level, as happens in the second paragraph. In the vast majority of cases where multiple events are described together at the clause level, this takes place in the headline paragraph, with the individual events then elaborated upon in the body of the document.

B.2 Preprocessing

```
header(txt99, [c#[place#[lima], punc#comma], c#[25, oct, 89, punc#openpar, efe,
punc#closepar, punc#period]]).
text(txt99, 1, 0, [c#[rsa#[police, have, reported, that], terrorists, tonight,
bombed, the, embassies, of, the, prc, and, the, soviet, place#[union],
punc#period]]).
text(txt99, 2, 0, [c#[the, bombs, caused, damage], c#[but, no, injuries,
punc#period]]).
text(txt99, 3, 1, [c#[a, car, bomb, exploded, in, front, of, the, prc, embassy,
punc#comma], c#[which, is, in, the, place#[lima], residential, district, of,
place#[san, isidro], punc#period]]).
text(txt99, 4, 0, [c#[meanwhile, punc#comma], c#[two, bombs, were, thrown, at,
a, ussr, embassy, vehicle, that, was, parked, in, front, of, the, embassy,
located, in, orrantia, district, punc#comma], c#[near, place#[san, isidro],
punc#period]]).
text(txt99, 5, 1, [c#[rsa#[police, said, the, attacks, were, carried, out,
almost, simultaneously, and, that], the, bombs, broke, windows, and, destroyed,
the, two, vehicles, punc#period]]).
text(txt99, 6, 1, [c#[no, one, has, claimed, responsibility, for, the, attacks,
so, far, punc#period]]).
text(txt99, 7, 0, [c#[police, sources, punc#comma], c#[however, punc#comma],
c#[have, said, the, attacks, could, have, been, carried, out, by, the, maoist,
shining, path, group, or, the, guevarist, tupac, amaru, revolutionary, movement,
punc#openpar, mrta, punc#closepar, group, punc#period]]).
text(txt99, 8, 0, [c#[rsa#[the, sources, also, said, that], the, shining,
path, has, attacked, soviet, interests, in, place#[peru], in, the, past,
punc#period]]).
text(txt99, 9, 1, [c#[in, july, 1989, the, shining, path, bombed, a, bus,
carrying, nearly, 50, soviet, marines, into, the, port, of, place#[el, callao],
punc#period]]).
text(txt99, 10, 0, [c#[fifteen, soviet, marines, were, wounded, punc#period]]).
text(txt99, 11, 1, [c#[some, 3, years, ago, two, marines, died, following, a,
shining, path, bombing, of, a, market, used, by, soviet, marines, punc#period]]).
text(txt99, 12, 1, [c#[in, another, incident, 3, years, ago, punc#comma], c#[a,
shining, path, militant, was, killed, by, soviet, embassy, guards, inside, the,
embassy, compound, punc#period]]).
text(txt99, 13, 0, [c#[the, terrorist, was, carrying, dynamite, punc#period]]).
text(txt99, 14, 1, [c#[the, attacks, today, come, after, shining, path, attacks,
during], c#[which, least, 10, buses, were, burned, throughout, place#[lima], on,
24, oct, punc#period]]).
```

This document is a good illustration of both the strengths and the weaknesses of the preprocessor. Reported speech and locations are successfully marked up, although one of the marked up reported speech fragments, sentence 5, is marginally incorrect due to the lack of a complementiser. Sentence 14 contains a good example of the sometimes bizarre clause segmentation performed by the preprocessor. In this case, clause segmentation has been triggered by the word *which*.

B.3 TAM output

The TAM identifies seven temporal phrases in the document. These are interpreted to give the forms shown below. Note the interpretations of the phrases *3 years ago*, *today* and *tonight* relative to the document date. As is to be expected for this eventful document, a large number of constraints are output. It should be noticed that, although the two events containing the *3 years ago* fragments above (sentences 11 and 12) are clearly describing two different events, so far as the TAM is concerned, there is nothing incompatible about the two interpreted dates — and consequently no constraint output for those sentences.

Temporal phrases identified

```
nl(txt99,ta,[0,1],[]).
nl(txt99,ta,[0,2],[25,oct,89]).
nl(txt99,ta,[1,1],[tonight]).
nl(txt99,ta,[2,1],[]).
nl(txt99,ta,[2,2],[]).
nl(txt99,ta,[3,1],[]).
nl(txt99,ta,[3,2],[]).
nl(txt99,ta,[4,1],[]).
nl(txt99,ta,[4,2],[]).
nl(txt99,ta,[4,3],[]).
nl(txt99,ta,[5,1],[]).
nl(txt99,ta,[6,1],[]).
nl(txt99,ta,[7,1],[]).
nl(txt99,ta,[7,2],[]).
nl(txt99,ta,[7,3],[]).
nl(txt99,ta,[8,1],[]).
nl(txt99,ta,[9,1],[in,july,1989]).
nl(txt99,ta,[10,1],[]).
nl(txt99,ta,[11,1],[3,years,ago]).
nl(txt99,ta,[12,1],[3,years,ago]).
nl(txt99,ta,[12,2],[]).
nl(txt99,ta,[13,1],[]).
nl(txt99,ta,[14,1],[today]).
nl(txt99,ta,[14,2],[on,24,oct]).
```

Interpreted phrases

```
-txt99-
  [[],[dd:25,mm:10,yy:1989,day:wed,abs:[wk:148,day:1029],idx:[s:0,c:2|_7120]|_6883]]
  [[dd:25,mm:10,yy:1989,abs:[wk:148,day:1029|_12890],daypart:night,rel:doc,
idx:[s:1,c:1|_14377]|_14124]]
  [[],[]]
  [[],[]]
  [[],[],[]]
  [[]]
  [[]]
  [[],[],[]]
  [[]]
  [[mm:7,yy:1989,idx:[s:9,c:1|_43707]|_43518]]
  [[]]
  [[yy:1986,rel:doc,idx:[s:11,c:1|_52559]|_52370]]
  [[yy:1986,rel:doc,idx:[s:12,c:1|_59380]|_59191],[]]
  [[]]
  [[dd:25,mm:10,yy:1989,abs:[wk:148,day:1029|_67999],rel:doc,idx:[s:14,c:1|_69269]|_69032],
   [dd:24,mm:10,yy:1989,day:tue,abs:[wk:148,day:1028],idx:[s:14,c:2|_77205]|_76968]]
```

Constraints output

```
constraint(txt99,time,[0,2],[9,1],1).
constraint(txt99,time,[0,2],[11,1],1).
constraint(txt99,time,[0,2],[12,1],1).
constraint(txt99,time,[0,2],[14,2],1).
constraint(txt99,time,[1,1],[9,1],1).
constraint(txt99,time,[1,1],[11,1],1).
constraint(txt99,time,[1,1],[12,1],1).
constraint(txt99,time,[1,1],[14,2],1).
constraint(txt99,time,[9,1],[11,1],1).
constraint(txt99,time,[9,1],[12,1],1).
constraint(txt99,time,[9,1],[14,1],1).
constraint(txt99,time,[9,1],[14,2],1).
constraint(txt99,time,[11,1],[14,1],1).
constraint(txt99,time,[11,1],[14,2],1).
constraint(txt99,time,[12,1],[14,1],1).
constraint(txt99,time,[12,1],[14,2],1).
constraint(txt99,time,[14,1],[14,2],1).
```

B.4 LAM output

Five locative phrases are identified by the LAM. As it turns out, most of these are compatible — hence the low number of constraints — with the exception of *the port of El Callao*, which is deemed to be incompatible with a city. It's not clear whether or not this is actually true in this case, as the gazetteer has no record within it of a port called *El Callao*. However, the typology of locations dictates that ports and cities are disjoint entities. This is simply a legacy of the structure implicit in the gazetteer.

Locative phrases identified

```
nl(txt99,la,[0,1],[place#[lima]]).
nl(txt99,la,[0,2],[]).
nl(txt99,la,[1,1],[]).
nl(txt99,la,[2,1],[]).
nl(txt99,la,[2,2],[]).
nl(txt99,la,[3,1],[]).
nl(txt99,la,[3,2],[place#[lima],residential,district,of,place#[san,isidro]]).
nl(txt99,la,[4,1],[]).
nl(txt99,la,[4,2],[]).
nl(txt99,la,[4,3],[]).
nl(txt99,la,[5,1],[]).
nl(txt99,la,[6,1],[]).
nl(txt99,la,[7,1],[]).
nl(txt99,la,[7,2],[]).
nl(txt99,la,[7,3],[]).
nl(txt99,la,[8,1],[in,place#[peru]]).
nl(txt99,la,[9,1],[the,port,of,place#[el,callao]]).
nl(txt99,la,[10,1],[]).
nl(txt99,la,[11,1],[]).
nl(txt99,la,[12,1],[]).
nl(txt99,la,[12,2],[]).
nl(txt99,la,[13,1],[]).
nl(txt99,la,[14,1],[]).
nl(txt99,la,[14,2],[throughout,place#[lima]]).
```

Interpreted phrases

```
-txt99-
 [[city:[lima],size:[1],airport:'n/a',port:'n/a',island:'n/a',province:[lima],
 country:[peru],lookup:yes,idx:[s:0,c:1|_4246]|_3961],[]]
 [[]]
 [[],[]]
 [[],[city:[lima],size:[1],airport:'n/a',port:'n/a',island:'n/a',province:[lima],
 country:[peru],lookup:yes,suburb:[san,isidro],idx:[s:3,c:2|_20417]|_20116]]
 [[],[],[]]
 [[]]
 [[]]
 [[],[],[]]
 [[country:[peru],lookup:yes,idx:[s:8,c:1|_38627]|_38438]]
 [[city:'n/a',airport:'n/a',port:[el,callao],island:'n/a',province:'n/a',
 country:[peru],lookup:no,idx:[s:9,c:1|_44952]|_44683]]
 [[]]
 [[]]
 [[],[]]
 [[]]
 [[],[city:[lima],size:[1],airport:'n/a',port:'n/a',island:'n/a',province:[lima],
 country:[peru],lookup:yes,idx:[s:14,c:2|_60904]|_60619]]
```

Constraints output

```
constraint(txt99,loc,[0,1],[9,1],1).
constraint(txt99,loc,[3,2],[9,1],1).
constraint(txt99,loc,[9,1],[14,2],1).
```

B.5 CPAM output

This document has two cue phrase, *meanwhile* and *in another incident.* The CPAM therefore proposes constraints between the clauses that each of these appear in and all those preceding them.

Cue phrases identified

```
nl(txt99,cpa,[4,1],[meanwhile]).
nl(txt99,cpa,[12,1],[in,another,incident]).
```

Constraints output

```
constraint(txt99,cue,[4,1],[0,_],1).
constraint(txt99,cue,[4,1],[1,_],1).
constraint(txt99,cue,[4,1],[2,_],1).
constraint(txt99,cue,[4,1],[3,_],1).
constraint(txt99,cue,[12,1],[0,_],1).
constraint(txt99,cue,[12,1],[1,_],1).
constraint(txt99,cue,[12,1],[2,_],1).
constraint(txt99,cue,[12,1],[3,_],1).
constraint(txt99,cue,[12,1],[4,_],1).
constraint(txt99,cue,[12,1],[5,_],1).
constraint(txt99,cue,[12,1],[6,_],1).
constraint(txt99,cue,[12,1],[7,_],1).
constraint(txt99,cue,[12,1],[8,_],1).
constraint(txt99,cue,[12,1],[9,_],1).
constraint(txt99,cue,[12,1],[10,_],1).
constraint(txt99,cue,[12,1],[11,_],1).
```

B.6 EM output

Figure B.1 shows graphically the constraint relationships that exist between the clauses in the text, with the constraints proposed by each analysis module in different line styles. This visualisation technique [Senator *et al.* 95] is often used in link-analysis tasks [Andrews & Peterson 90] to present complex data.

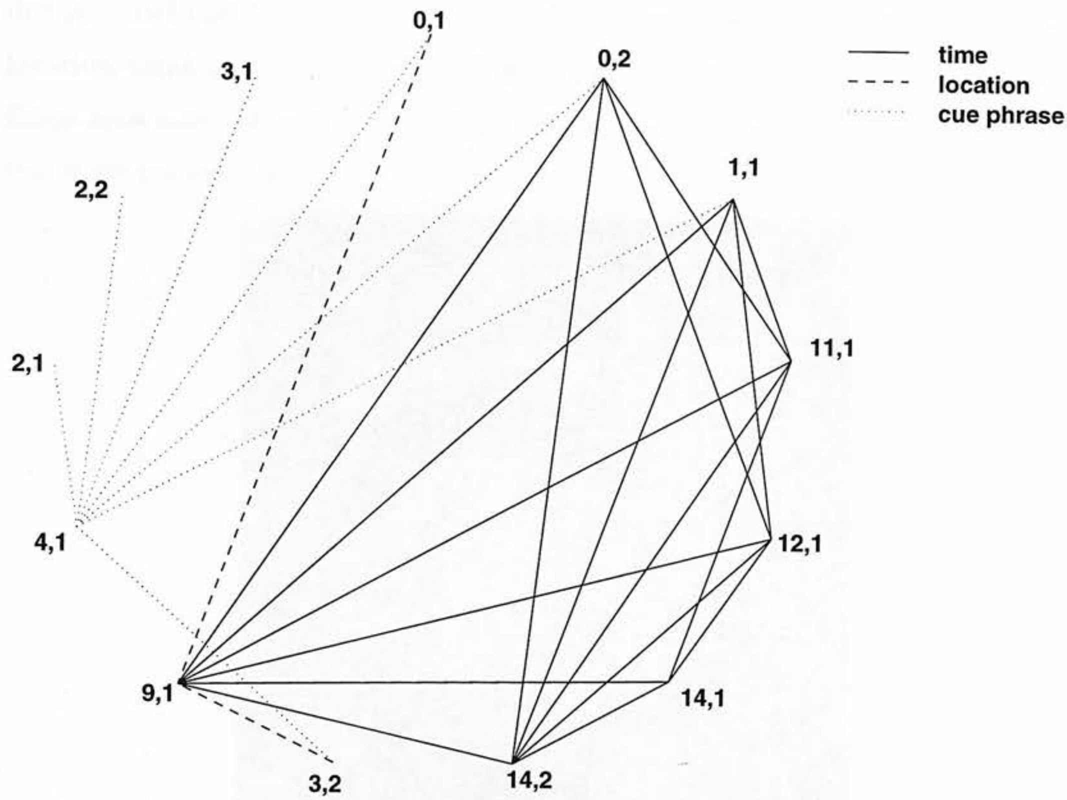


Figure B.1: Wagon-wheel visualisation of constraints

The 128 grids output by CONTESS's configurations are represented unidimensionally in figure B.2. Each horizontal strip represents one grid, with changes in intensity along the strip denoting event shifts. Therefore, strips with uniform intensity indicate uneventful interpretations, whereas strips with many intensity changes indicate eventful representations. Strips should be read from left to right in the same way that clause-event grids are read from top to bottom.

At the right of the image is a broken strip of seven parts. This shows the configuration that generated each 2-dimensional grid. The strip components are, from left to right, **time**, **location**, **strategy**, **sentence** gelling, **clause** gelling, **frequency** analysis and **cue** phrases. Bright areas indicate that the appropriate component was selected, while darker areas indicate that it was not selected.

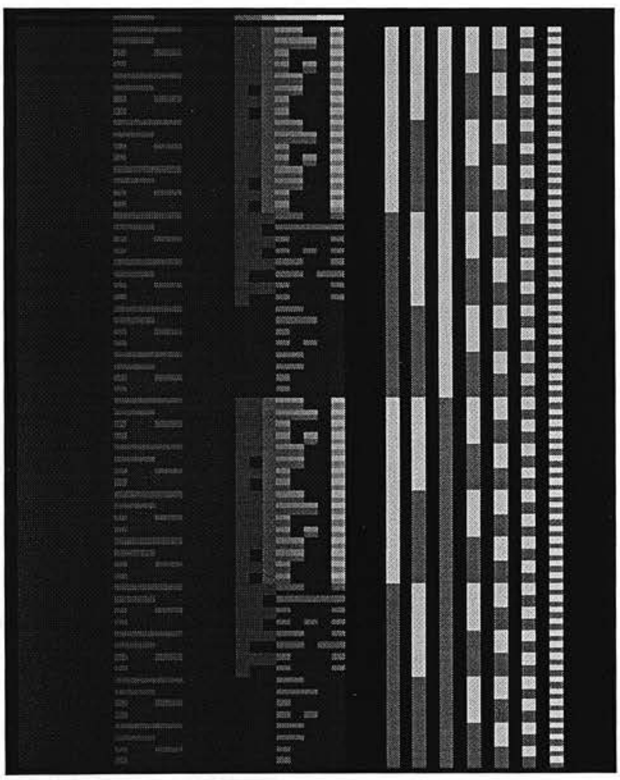


Figure B.2: unidimensional visualisation of each configuration's grid for detailed example

Appendix C

Results tables

This appendix contains the full results tables (documents, configurations and components) for the example corpus, the three development corpus tests (original grammar, modified grammar and modified grammar with CRaP filter activated) and for the test corpus. As an aid to the reader, the structure of each section is outlined below, with actual tables highlighted in bold font. Document analysis tables are split over two pages (except in the case of the example corpus, which contains only one document), and configuration analysis tables split over three pages.

- Documents
  - **Numeric document analysis**
  - **Ranked document analysis**
- Configurations
  - **Numeric configuration analysis**
  - **Ranked configuration analysis**
- Components
  - **Numeric component analysis**
  - **Ranked component analysis**

The **TLSPCFU** notation for system configuration is used in the following tables. As a reminder, these stand for: **T**ime analysis, **L**ocation analysis, clustering **S**trategy, gelling of sentences within **P**aragraphs, **C**lause gelling, **F**requency analysis and c**U**e phrase analysis.

C.1 Example corpus results

Documents

Numeric document analysis

DOCUMENTS
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
00	34.5	100.0	100.0	100.0	100.0	100.0	92.7	92.7	87.3	87.3	87.3	87.3			34.5	61.8
AVE	34.5	100.0	100.0	100.0	100.0	100.0	92.7	92.7	87.3	87.3	87.3	87.3			34.5	61.8

KEY:
1: s1101111 2: s1101100 3: s1101110 4: s1101101 5: s1100111 6: s1100101 7: s1001101
8: s1001111 9: s1101001 10: s1101011 C:

Ranked document analysis

DOCUMENTS
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile.ranked

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
00	1.0	18.0	1.0	1.0	1.0	1.0	2.0	2.0	3.0	3.0	3.0	3.0				
AVE	-	-	1.0	1.0	1.0	1.0	2.0	2.0	3.0	3.0	3.0	3.0			-	-

KEY:
1: s1101111 2: s1101100 3: s1101110 4: s1101101 5: s1100111 6: s1100101 7: s1001101
8: s1001111 9: s1101001 10: s1101011 C:

Configurations

Numeric configuration analysis

CONFIGURATIONS																	
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
1	s1101111	100.0	100.0	0.0	100.0	0	0	0	0	0	0	0	0	0	0	1	0.00
2	s1101100	100.0	100.0	0.0	100.0	0	0	0	0	0	0	0	0	0	0	1	0.00
3	s1101110	100.0	100.0	0.0	100.0	0	0	0	0	0	0	0	0	0	0	1	0.00
4	s1101101	100.0	100.0	0.0	100.0	0	0	0	0	0	0	0	0	0	0	1	0.00
5	s1100111	92.7	92.7	0.0	92.7	0	0	0	0	0	0	0	0	0	1	0	0.00
6	s1100101	92.7	92.7	0.0	92.7	0	0	0	0	0	0	0	0	0	1	0	0.00
7	s1001101	87.3	87.3	0.0	87.3	0	0	0	0	0	0	0	0	1	0	0	0.00
8	s1001111	87.3	87.3	0.0	87.3	0	0	0	0	0	0	0	0	1	0	0	0.00
9	s1101001	87.3	87.3	0.0	87.3	0	0	0	0	0	0	0	0	1	0	0	0.00
10	s1101011	87.3	87.3	0.0	87.3	0	0	0	0	0	0	0	0	1	0	0	0.00
11	s0101101	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
12	s0101110	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
13	s0101111	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
14	s1000111	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
15	s1000101	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
16	s0101001	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
17	s0101100	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
18	s0101011	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
19	s0011101	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
20	s0001111	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
21	s1100001	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
22	s1100011	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
23	s0001101	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
24	s0011111	78.2	78.2	0.0	78.2	0	0	0	0	0	0	0	1	0	0	0	0.00
25	s1101000	72.7	72.7	0.0	72.7	0	0	0	0	0	0	0	1	0	0	0	0.00
26	s1101010	72.7	72.7	0.0	72.7	0	0	0	0	0	0	0	1	0	0	0	0.00
27	s1001110	70.9	70.9	0.0	70.9	0	0	0	0	0	0	0	1	0	0	0	0.00
28	s1001100	70.9	70.9	0.0	70.9	0	0	0	0	0	0	0	1	0	0	0	0.00
29	s1011110	70.9	70.9	0.0	70.9	0	0	0	0	0	0	0	1	0	0	0	0.00
30	s1011100	70.9	70.9	0.0	70.9	0	0	0	0	0	0	0	1	0	0	0	0.00
31	s1111100	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
32	s1111101	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
33	s0111111	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
34	s0111110	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
35	s0111101	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
36	s0111100	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
37	s0111011	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
38	s1111110	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
39	s0111001	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
40	s1111111	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
41	s1001011	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
42	s1100110	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
43	s1100100	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
44	s1001001	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
45	s0100111	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
46	s0100101	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
47	s0100011	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
48	s0100001	67.3	67.3	0.0	67.3	0	0	0	0	0	0	1	0	0	0	0	0.00
49	s1111011	65.5	65.5	0.0	65.5	0	0	0	0	0	0	1	0	0	0	0	0.00
50	s1111001	65.5	65.5	0.0	65.5	0	0	0	0	0	0	1	0	0	0	0	0.00
continued below																	

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile

NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
51	s1110101	65.5	65.5	0.0	65.5	0	0	0	0	0	0	1	0	0	0	0	0.00
52	s1011101	65.5	65.5	0.0	65.5	0	0	0	0	0	0	1	0	0	0	0	0.00
53	s1110111	65.5	65.5	0.0	65.5	0	0	0	0	0	0	1	0	0	0	0	0.00
54	s1011111	65.5	65.5	0.0	65.5	0	0	0	0	0	0	1	0	0	0	0	0.00
55	s0001001	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
56	s0101000	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
57	s0011001	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
58	s0010111	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
59	s0010101	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
60	s0101010	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
61	s0011011	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
62	s0001011	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
63	s1000001	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
64	s1000011	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
65	s1110011	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
66	s0000111	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
67	s1110001	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
68	s1010101	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
69	s0000101	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
70	s1010111	60.0	60.0	0.0	60.0	0	0	0	0	0	0	1	0	0	0	0	0.00
71	s0110011	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
72	s0110001	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
73	s0110101	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
74	s0110111	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
75	s1011001	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
76	s1111000	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
77	s1011011	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
78	s1111010	58.2	58.2	0.0	58.2	0	0	0	0	0	1	0	0	0	0	0	0.00
79	s1011000	56.4	56.4	0.0	56.4	0	0	0	0	0	1	0	0	0	0	0	0.00
80	s1011010	56.4	56.4	0.0	56.4	0	0	0	0	0	1	0	0	0	0	0	0.00
81	s0100110	56.4	56.4	0.0	56.4	0	0	0	0	0	1	0	0	0	0	0	0.00
82	s0100100	56.4	56.4	0.0	56.4	0	0	0	0	0	1	0	0	0	0	0	0.00
83	s1001000	56.4	56.4	0.0	56.4	0	0	0	0	0	1	0	0	0	0	0	0.00
84	s1001010	56.4	56.4	0.0	56.4	0	0	0	0	0	1	0	0	0	0	0	0.00
85	s1010011	52.7	52.7	0.0	52.7	0	0	0	0	0	1	0	0	0	0	0	0.00
86	s0111000	52.7	52.7	0.0	52.7	0	0	0	0	0	1	0	0	0	0	0	0.00
87	s1010001	52.7	52.7	0.0	52.7	0	0	0	0	0	1	0	0	0	0	0	0.00
88	s0111010	52.7	52.7	0.0	52.7	0	0	0	0	0	1	0	0	0	0	0	0.00
89	s0110110	50.9	50.9	0.0	50.9	0	0	0	0	0	1	0	0	0	0	0	0.00
90	s1110110	50.9	50.9	0.0	50.9	0	0	0	0	0	1	0	0	0	0	0	0.00
91	s1110100	50.9	50.9	0.0	50.9	0	0	0	0	0	1	0	0	0	0	0	0.00
92	s1100000	50.9	50.9	0.0	50.9	0	0	0	0	0	1	0	0	0	0	0	0.00
93	s0110100	50.9	50.9	0.0	50.9	0	0	0	0	0	1	0	0	0	0	0	0.00
94	s1100010	50.9	50.9	0.0	50.9	0	0	0	0	0	1	0	0	0	0	0	0.00
95	s1000100	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
96	s1000110	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
97	s0010001	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
98	s1010100	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
99	s1010110	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
100	s0010011	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00

continued below

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
101	s0000001	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
102	s0100000	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
103	s0000011	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
104	s0100010	45.5	45.5	0.0	45.5	0	0	0	0	1	0	0	0	0	0	0	0.00
105	s1110010	43.6	43.6	0.0	43.6	0	0	0	0	1	0	0	0	0	0	0	0.00
106	s1110000	43.6	43.6	0.0	43.6	0	0	0	0	1	0	0	0	0	0	0	0.00
107	s0110010	41.8	41.8	0.0	41.8	0	0	0	0	1	0	0	0	0	0	0	0.00
108	s0110000	41.8	41.8	0.0	41.8	0	0	0	0	1	0	0	0	0	0	0	0.00
109	s1000010	38.2	38.2	0.0	38.2	0	0	0	1	0	0	0	0	0	0	0	0.00
110	s1000000	38.2	38.2	0.0	38.2	0	0	0	1	0	0	0	0	0	0	0	0.00
111	s1010000	38.2	38.2	0.0	38.2	0	0	0	1	0	0	0	0	0	0	0	0.00
112	s1010010	38.2	38.2	0.0	38.2	0	0	0	1	0	0	0	0	0	0	0	0.00
113	s0010000	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
114	s0011110	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
115	s0011010	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
116	s0011100	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
117	s0001100	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
118	s0001010	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
119	s0010110	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
120	s0001000	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
121	s0011000	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
122	s0000110	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
123	s0010010	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
124	s0000100	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
125	s0001110	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
126	s0000010	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
127	s0010100	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
128	s0000000	34.5	34.5	0.0	34.5	0	0	0	1	0	0	0	0	0	0	0	0.00
	AVE	59.9	59.9	0.0	59.9	0.0	0.0	0.0	0.2	0.1	0.2	0.3	0.2	0.0	0.0	0.0	0.00

Ranked configuration analysis

CONFIGURATIONS
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
1	s1101111	1	1	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0.00
2	s1101100	1	1	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0.00
3	s1101110	1	1	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0.00
4	s1101101	1	1	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0.00
5	s1100111	2	2	0	2.0	0	0	0	0	0	0	0	0	0	0	0	0.00
6	s1100101	2	2	0	2.0	0	0	0	0	0	0	0	0	0	0	0	0.00
7	s1001101	3	3	0	3.0	0	0	0	0	0	0	0	0	0	0	0	0.00
8	s1001111	3	3	0	3.0	0	0	0	0	0	0	0	0	0	0	0	0.00
9	s1101001	3	3	0	3.0	0	0	0	0	0	0	0	0	0	0	0	0.00
10	s1101011	3	3	0	3.0	0	0	0	0	0	0	0	0	0	0	0	0.00
11	s0101101	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
12	s0101110	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
13	s0101111	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
14	s1000111	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
15	s1000101	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
16	s0101001	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
17	s0101100	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
18	s0101011	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
19	s0011101	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
20	s0001111	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
21	s1100001	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
22	s1100011	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
23	s0001101	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
24	s0011111	4	4	0	4.0	0	0	0	0	0	0	0	0	0	0	0	0.00
25	s1101000	5	5	0	5.0	0	0	0	0	0	0	0	0	0	0	0	0.00
26	s1101010	5	5	0	5.0	0	0	0	0	0	0	0	0	0	0	0	0.00
27	s1001110	6	6	0	6.0	0	0	0	0	0	0	0	0	0	0	0	0.00
28	s1001100	6	6	0	6.0	0	0	0	0	0	0	0	0	0	0	0	0.00
29	s1011110	6	6	0	6.0	0	0	0	0	0	0	0	0	0	0	0	0.00
30	s1011100	6	6	0	6.0	0	0	0	0	0	0	0	0	0	0	0	0.00
31	s1111100	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
32	s1111101	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
33	s0111111	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
34	s0111110	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
35	s0111101	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
36	s0111100	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
37	s0111011	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
38	s1111110	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
39	s0111001	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
40	s1111111	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
41	s1001011	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
42	s1100110	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
43	s1100100	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
44	s1001001	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
45	s0100111	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
46	s0100101	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
47	s0100011	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
48	s0100001	7	7	0	7.0	0	0	0	0	0	0	0	0	0	0	0	0.00
49	s1111011	8	8	0	8.0	0	0	0	0	0	0	0	0	0	0	0	0.00
50	s1111001	8	8	0	8.0	0	0	0	0	0	0	0	0	0	0	0	0.00

continued below

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
51	s1110101	8	8	0	8.0	0	0	0	0	0	0	0	0	0	0	0	0.00
52	s1011101	8	8	0	8.0	0	0	0	0	0	0	0	0	0	0	0	0.00
53	s1110111	8	8	0	8.0	0	0	0	0	0	0	0	0	0	0	0	0.00
54	s1011111	8	8	0	8.0	0	0	0	0	0	0	0	0	0	0	0	0.00
55	s0001001	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
56	s0101000	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
57	s0011001	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
58	s0010111	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
59	s0010101	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
60	s0101010	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
61	s0011011	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
62	s0001011	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
63	s1000001	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
64	s1000011	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
65	s1110011	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
66	s0000111	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
67	s1110001	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
68	s1010101	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
69	s0000101	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
70	s1010111	9	9	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0.00
71	s0110011	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
72	s0110001	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
73	s0110101	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
74	s0110111	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
75	s1011001	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
76	s1111000	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
77	s1011011	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
78	s1111010	10	10	0	10.0	0	0	0	0	0	0	0	0	0	0	0	0.00
79	s1011000	11	11	0	11.0	0	0	0	0	0	0	0	0	0	0	0	0.00
80	s1011010	11	11	0	11.0	0	0	0	0	0	0	0	0	0	0	0	0.00
81	s0100110	11	11	0	11.0	0	0	0	0	0	0	0	0	0	0	0	0.00
82	s0100100	11	11	0	11.0	0	0	0	0	0	0	0	0	0	0	0	0.00
83	s1001000	11	11	0	11.0	0	0	0	0	0	0	0	0	0	0	0	0.00
84	s1001010	11	11	0	11.0	0	0	0	0	0	0	0	0	0	0	0	0.00
85	s1010011	12	12	0	12.0	0	0	0	0	0	0	0	0	0	0	0	0.00
86	s0111000	12	12	0	12.0	0	0	0	0	0	0	0	0	0	0	0	0.00
87	s1010001	12	12	0	12.0	0	0	0	0	0	0	0	0	0	0	0	0.00
88	s0111010	12	12	0	12.0	0	0	0	0	0	0	0	0	0	0	0	0.00
89	s0110110	13	13	0	13.0	0	0	0	0	0	0	0	0	0	0	0	0.00
90	s1110110	13	13	0	13.0	0	0	0	0	0	0	0	0	0	0	0	0.00
91	s1110100	13	13	0	13.0	0	0	0	0	0	0	0	0	0	0	0	0.00
92	s1100000	13	13	0	13.0	0	0	0	0	0	0	0	0	0	0	0	0.00
93	s0110100	13	13	0	13.0	0	0	0	0	0	0	0	0	0	0	0	0.00
94	s1100010	13	13	0	13.0	0	0	0	0	0	0	0	0	0	0	0	0.00
95	s1000100	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
96	s1000110	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
97	s0010001	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
98	s1010100	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
99	s1010110	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
100	s0010011	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00

continued below

CONFIGURATIONS (continued)																	
/hame/jeremyc/phd/eval/sys/scores/DEMO/scorefile.ranked																	
NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
101	s0000001	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
102	s0100000	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
103	s0000011	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
104	s0100010	14	14	0	14.0	0	0	0	0	0	0	0	0	0	0	0	0.00
105	s1110010	15	15	0	15.0	0	0	0	0	0	0	0	0	0	0	0	0.00
106	s1110000	15	15	0	15.0	0	0	0	0	0	0	0	0	0	0	0	0.00
107	s0110010	16	16	0	16.0	0	0	0	0	0	0	0	0	0	0	0	0.00
108	s0110000	16	16	0	16.0	0	0	0	0	0	0	0	0	0	0	0	0.00
109	s1000010	17	17	0	17.0	0	0	0	0	0	0	0	0	0	0	0	0.00
110	s1000000	17	17	0	17.0	0	0	0	0	0	0	0	0	0	0	0	0.00
111	s1010000	17	17	0	17.0	0	0	0	0	0	0	0	0	0	0	0	0.00
112	s1010010	17	17	0	17.0	0	0	0	0	0	0	0	0	0	0	0	0.00
113	s0010000	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
114	s0011110	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
115	s0011010	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
116	s0011100	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
117	s0001100	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
118	s0001010	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
119	s0010110	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
120	s0001000	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
121	s0011000	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
122	s0000110	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
123	s0010010	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
124	s0000100	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
125	s0001110	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
126	s0000010	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
127	s0010100	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
128	s0000000	18	18	0	18.0	0	0	0	0	0	0	0	0	0	0	0	0.00
	AVE	9.8	9.8	0.0	9.8	0	0	0	0	0	0	0	0	0	0	0	0.00

Components

Numeric (left) and ranked (right) component analysis

sT------	9.999987	sT------	2.718750
s-L-----	12.727263	s-L-----	3.718750
s--S----	-9.204544	s--S----	-2.468750
s---P---	12.272725	s---P---	3.656250
s----C--	9.659087	s----C--	2.718750
s-----F-	0.000000	s-----F-	0.000000
s------U	16.022706	s------U	5.156250

C.2 Development corpus results I

This section contains the document, configuration and component analysis tables (numerical and ranked) for the development corpus using the original TAM and LAM grammars, i.e. before rule evaluation.

The component analysis tables were first presented in section 9.2, but are repeated here for purposes of completeness.

Documents

Numeric document analysis

DOCUMENTS																
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
00	57.9	96.7	64.8	64.7	82.5	80.6	62.1	62.1	79.3	78.7	60.7	60.7		s01-0101/2	60.0	44.6
01	53.7	79.2	58.4	58.4	58.4	58.4	58.4	58.4	58.4	58.4	57.6	57.6		s11--10-/8	53.7	55.0
02	38.9	92.1	66.8	66.8	66.8	66.8	43.7	43.7	47.9	47.9	92.1	92.1			41.1	57.9
03	66.8	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	66.8	66.8			78.1	27.5
04	48.4	90.2	56.2	56.2	77.8	77.8	56.2	56.2	77.8	77.8	90.2	90.2			69.3	37.3
05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.2
06	46.2	99.1	99.1	99.1	91.7	91.7	99.1	99.1	91.7	91.7	64.6	64.6			72.9	37.2
07	47.1	100.0	88.2	88.2	88.2	88.2	88.2	88.2	88.2	88.2	47.1	47.1		s00-----/32	100.0	21.3
08	85.0	100.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	88.5	88.5		s1---0--/32	96.0	15.8
09	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	32.7
10	73.5	100.0	73.5	73.5	77.0	77.0	73.5	73.5	77.0	77.0	80.4	80.4		s0------/64	100.0	13.6
11	25.7	100.0	54.9	54.9	59.2	59.2	54.9	54.9	59.2	59.2	29.2	30.1		s00-----/32	100.0	19.4
12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.3
13	51.0	85.4	85.4	85.4	70.3	70.3	85.4	85.4	70.3	70.3	56.9	58.5			82.6	25.9
14	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0			87.0	24.7
15	89.5	100.0	94.9	94.9	97.4	97.4	94.9	94.9	97.4	97.4	94.7	94.7		s01--1--/16	89.5	20.3
16	55.1	100.0	89.7	89.7	55.1	55.1	89.7	89.7	55.1	55.1	58.3	56.6		s0------/64	100.0	6.9
17	42.6	100.0	96.4	96.4	86.3	86.3	96.4	96.4	86.3	86.3	100.0	100.0			96.4	16.6
18	50.0	64.0	60.3	60.3	64.0	64.0	60.3	60.3	64.0	64.0	57.4	57.4			51.5	47.1
19	85.4	100.0	96.9	96.9	85.4	85.4	96.9	96.9	85.4	85.4	85.4	85.4		s0------/64	100.0	11.6
20	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.0
21	70.8	100.0	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2		s00-----/32	100.0	15.3
22	41.2	90.3	41.2	41.2	41.2	41.2	41.2	41.2	41.2	41.2	82.6	82.6		s0------/64	90.3	19.9
23	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	10.4
24	36.9	68.0	62.6	62.6	68.0	68.0	62.6	62.6	68.0	68.0	58.6	58.6			36.9	60.8
25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	18.2
26	47.5	67.1	63.8	63.6	67.1	66.9	63.8	63.6	67.1	66.9	67.1	67.1			54.5	49.0
27	47.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	17.6
28	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	27.3
29	37.5	88.0	78.6	78.6	47.4	47.4	78.6	78.6	47.4	47.4	83.2	83.2		s01-0-0-/8	75.2	33.8
30	73.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	10.1
31	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	23.6
32	34.8	69.3	59.5	59.5	69.3	69.3	59.5	59.5	69.3	69.3	60.4	60.4			34.8	61.1
33	37.6	92.4	70.0	70.0	84.3	84.3	70.0	70.0	84.3	84.3	56.7	56.7		s01-110-/4	37.6	57.6
34	76.1	83.0	79.7	79.7	79.7	79.7	79.7	79.7	79.7	79.7	83.0	83.0			78.2	27.4
35	45.0	62.6	45.0	45.0	62.6	62.6	45.0	45.0	62.6	62.6	62.6	62.6			54.4	43.9
36	56.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	65.5
37	53.3	100.0	100.0	100.0	53.3	53.3	100.0	100.0	53.3	53.3	90.5	90.5			74.3	34.8
38	28.6	73.7	61.6	61.6	73.7	73.7	61.6	61.6	73.7	73.7	47.8	47.8			28.6	66.2
39	47.3	66.4	60.2	60.2	66.4	66.4	60.2	60.2	66.4	66.4	51.5	51.5			47.3	56.6
40	20.0	98.2	87.1	87.1	87.1	87.1	87.1	87.1	87.1	87.1	98.2	98.2			20.0	59.7
41	54.1	100.0	81.4	81.4	100.0	100.0	54.1	54.1	54.1	54.1	91.9	91.9			54.1	46.3
42	36.4	94.1	94.1	94.1	72.1	72.1	94.1	94.1	72.1	72.1	91.1	91.1			90.9	15.8
43	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6			95.6	14.5
44	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	7.2
45	72.9	93.7	88.0	88.0	88.0	88.0	88.0	88.0	88.0	88.0	93.7	93.7			72.9	35.9
46	48.8	57.6	53.9	53.9	55.6	55.6	53.9	53.9	55.6	55.6	52.4	52.4		s-1-110-/8	48.8	51.7
47	48.2	84.1	65.6	65.6	77.2	77.2	65.6	65.6	77.2	77.2	48.2	48.2		s-------/72	84.1	34.1
48	35.9	100.0	82.1	82.1	82.1	82.1	35.9	35.9	35.9	35.9	82.1	82.1		s-1--1-1/16	35.9	57.7
49	89.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	50.9
continued below																

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
50	89.8	100.0	89.8	89.8	89.8	89.8	89.8	89.8	89.8	89.8	89.8	89.8		s-1--1--/32	89.8	30.6
51	84.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	33.3
52	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	7.7
53	50.8	88.1	84.9	84.9	88.1	88.1	61.8	61.8	61.8	61.8	51.8	50.8			61.8	40.7
54	53.8	97.3	97.3	97.3	84.7	84.7	97.3	97.3	84.7	84.7	54.8	53.8			97.3	11.0
55	67.8	100.0	82.0	82.0	82.0	82.0	100.0	100.0	100.0	100.0	67.8	67.8			100.0	16.0
56	32.3	78.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7		s-11-1--/16	32.3	62.4
57	52.9	75.7	64.3	64.3	64.3	64.3	64.3	64.3	64.3	64.3	55.2	55.2		s1---00-/16	60.0	41.0
58	48.8	100.0	91.1	91.1	83.1	83.1	91.1	91.1	83.1	83.1	95.5	95.5		s00-----/32	100.0	8.6
59	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.6
60	94.0	100.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0		s0------/64	100.0	10.6
61	69.0	100.0	100.0	100.0	99.0	99.0	100.0	100.0	99.0	99.0	93.6	93.6			69.0	36.9
62	37.5	64.1	59.5	54.8	59.5	54.8	42.5	42.5	42.5	42.5	64.1	64.1			37.5	59.2
63	51.8	98.4	92.9	92.9	98.4	98.4	92.9	92.9	98.4	98.4	54.2	54.2			62.8	38.7
64	51.5	88.9	88.9	88.9	51.5	51.5	88.9	88.9	51.5	51.5	81.6	81.6			80.5	24.4
65	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2			75.2	27.6
66	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.2
67	42.9	100.0	91.5	91.5	80.2	80.2	91.5	91.5	80.2	80.2	95.7	95.7		s00-----/32	100.0	8.6
68	53.3	100.0	100.0	100.0	81.9	81.9	100.0	100.0	81.9	81.9	74.3	74.3			90.5	25.2
69	41.6	73.2	66.4	66.4	73.2	73.2	66.4	66.4	73.2	73.2	49.3	49.3			64.1	46.7
70	53.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	10.1
71	49.4	94.7	74.5	74.5	78.0	78.0	74.5	74.5	78.0	78.0	94.7	94.7			58.5	45.0
72	46.2	84.6	84.6	84.6	84.6	84.6	84.6	84.6	84.6	84.6	46.2	46.2			53.8	44.9
73	25.3	66.7	54.9	54.9	54.9	54.9	54.9	54.9	54.9	54.9	51.3	51.3		s01-110-/4	25.3	68.3
74	44.9	100.0	73.5	73.5	64.7	64.7	100.0	100.0	89.0	89.0	73.5	73.5			77.2	42.6
75	54.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	93.9	93.9			83.0	24.6
76	46.2	84.6	64.0	64.0	70.2	70.2	64.0	64.0	70.2	70.2	80.9	80.9		s01--11-/8	56.0	45.2
77	89.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.9
78	24.6	92.9	83.1	83.1	92.9	92.9	71.1	71.1	70.8	70.8	60.0	60.0			24.6	67.7
79	40.5	100.0	82.8	82.8	82.6	82.6	82.8	82.8	82.6	82.6	86.3	86.3		s00-----/32	100.0	8.4
80	41.0	100.0	76.9	76.9	76.9	76.9	76.9	76.9	76.9	76.9	76.9	76.9		s11--11-/8	48.7	57.7
81	84.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	24.4
82	31.8	100.0	87.9	87.9	100.0	100.0	56.1	56.1	72.7	72.7	87.9	87.9			31.8	65.2
83	39.6	85.8	46.7	46.7	46.7	46.7	46.7	46.7	46.7	46.7	70.4	70.4		s0------/64	85.8	29.1
84	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	12.5
85	40.8	100.0	85.1	85.1	50.9	50.9	100.0	100.0	100.0	100.0	55.6	55.6			100.0	21.6
86	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.7
87	52.6	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	81.6	81.6			57.4	46.3
88	16.7	86.8	23.4	23.4	20.0	20.0	23.4	23.4	20.0	20.0	20.0	20.0		s111110-/2	16.7	77.3
89	24.2	98.8	42.5	42.5	98.8	98.8	28.1	28.1	71.0	71.0	54.8	53.0			24.2	72.5
90	47.3	100.0	63.7	63.7	63.7	63.7	63.7	63.7	63.7	63.7	100.0	100.0			47.3	57.1
91	46.1	74.6	70.1	70.1	65.0	65.0	70.1	70.1	65.0	65.0	53.9	55.2		s00-----/32	74.6	29.9
92	57.2	62.8	57.2	57.2	57.2	57.2	62.8	62.8	62.8	62.8	57.2	57.2			62.8	39.7
93	49.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	12.9
94	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	75.0
95	51.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	17.4
96	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.4
97	35.5	70.9	62.4	62.4	67.4	67.4	65.7	65.7	70.9	70.9	66.4	66.4			36.0	63.0
98	38.2	100.0	56.4	56.4	56.4	56.4	56.4	56.4	56.4	56.4	56.4	56.4		s-1-110-/8	38.2	50.9
99	31.2	92.8	82.2	82.2	92.8	92.8	65.2	65.2	78.6	78.6	76.8	76.8			31.2	63.0
AVE	60.0	92.1	82.2	82.1	81.7	81.6	80.7	80.7	80.2	80.2	79.0	78.9			75.5	34.1

KEY:
1: s1010101 2: s1000101 3: s1011101 4: s1001101 5: s1010100 6: s1000100 7: s1011100
8: s1001100 9: s1010111 10: s1011111 C:

Ranked document analysis

DOCUMENTS /hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile.ranked																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[|]	[?]
00	1.5	127.5	77.0	78.0	93.0	93.0	29.0	36.0	39.0	40.0	91.0	93.0		s01-0101/2		
01	4.5	112.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5		s11--10-/8		
02	5.5	127.5	42.5	42.5	101.5	101.5	42.5	42.5	77.5	77.5	96.5	96.5		s1---1-1/10		
03	8.5	120.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	24.5	24.5				
04	8.5	124.5	96.5	96.5	96.5	96.5	20.5	20.5	20.5	20.5	96.5	96.5		s1---11-/16		
05	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
06	2.5	124.5	2.5	2.5	2.5	2.5	6.5	6.5	6.5	6.5	14.5	14.5				
07	16.5	124.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5		s00-----/32		
08	16.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	16.5	16.5				
09	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
10	32.5	124.5	124.5	124.5	124.5	124.5	116.5	116.5	116.5	116.5	80.5	80.5		s0------/64		
11	16.5	127.5	73.5	73.5	73.5	73.5	66.5	66.5	66.5	66.5	40.5	40.5		s00-----/32		
12	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
13	4.5	124.5	4.5	4.5	4.5	4.5	92.5	92.5	92.5	92.5	12.5	12.5				
14	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
15	8.5	112.5	26.5	26.5	26.5	26.5	18.5	18.5	18.5	18.5	88.5	88.5		s01--1--/16		
16	32.5	124.5	76.5	76.5	76.5	76.5	124.5	124.5	124.5	124.5	68.5	68.5		s0------/64		
17	12.5	127.5	42.5	42.5	42.5	42.5	92.5	92.5	92.5	92.5	12.5	12.5				
18	6.5	120.5	14.5	14.5	14.5	14.5	6.5	6.5	6.5	6.5	64.5	64.5				
19	32.5	116.5	72.5	72.5	72.5	72.5	116.5	116.5	116.5	116.5	72.5	72.5		s0------/64		
20	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
21	16.5	126.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5		s00-----/32		
22	32.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	96.5	96.5		s0------/64		
23	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
24	4.5	112.5	12.5	12.5	12.5	12.5	4.5	4.5	4.5	4.5	42.5	42.5				
25	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
26	3.5	124.5	17.5	19.5	17.5	19.5	3.5	7.5	3.5	7.5	50.5	50.5				
27	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
28	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
29	4.5	127.5	34.5	34.5	34.5	34.5	122.5	122.5	122.5	122.5	26.5	26.5		s01-0-0-/8		
30	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
31	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
32	2.5	112.5	22.5	22.5	22.5	22.5	2.5	2.5	2.5	2.5	70.5	70.5				
33	2.5	112.5	14.5	14.5	14.5	14.5	6.5	6.5	6.5	6.5	78.5	78.5		s01-110-/4		
34	8.5	120.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	24.5	24.5		s1---11-/16		
35	12.5	124.5	124.5	124.5	124.5	124.5	12.5	12.5	12.5	12.5	116.5	116.5				
36	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
37	4.5	124.5	4.5	4.5	4.5	4.5	124.5	124.5	124.5	124.5	40.5	40.5				
38	4.5	96.5	12.5	12.5	12.5	12.5	4.5	4.5	4.5	4.5	48.5	48.5				
39	4.5	96.5	20.5	20.5	20.5	20.5	4.5	4.5	4.5	4.5	28.5	28.5				
40	4.5	112.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	64.5	64.5		s10--11-/8		
41	2.5	112.5	26.5	26.5	112.5	112.5	2.5	2.5	112.5	112.5	94.5	94.5				
42	2.5	127.5	2.5	2.5	2.5	2.5	110.5	110.5	110.5	110.5	18.5	18.5				
43	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
44	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
45	8.5	96.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	48.5	48.5		s1---11-/16		
46	4.5	112.5	22.5	22.5	22.5	22.5	14.5	14.5	14.5	14.5	84.5	84.5		s-1-110-/8		
47	36.5	120.5	92.5	92.5	92.5	92.5	80.5	80.5	80.5	80.5	80.5	80.5		s-------/72		
48	8.5	112.5	28.5	28.5	112.5	112.5	28.5	28.5	112.5	112.5	88.5	88.5		s-1--1-1/16		
49	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
continued below																

DOCUMENTS (continued)																
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile.ranked																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[!]	[?]
50	16.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5		s-1--1--/32		
51	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
52	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
53	4.5	124.5	12.5	12.5	64.5	64.5	4.5	4.5	64.5	64.5	28.5	28.5				
54	8.5	124.5	8.5	8.5	8.5	8.5	92.5	92.5	92.5	92.5	8.5	8.5				
55	32.5	120.5	104.5	104.5	32.5	32.5	104.5	104.5	32.5	32.5	72.5	72.5				
56	8.5	112.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	88.5	88.5		s-11-1--/16		
57	8.5	120.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	8.5	8.5				
58	16.5	124.5	58.5	58.5	58.5	58.5	62.5	62.5	62.5	62.5	44.5	44.5		s00-----/32		
59	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
60	32.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	80.5	80.5		s0------/64		
61	4.5	96.5	4.5	4.5	4.5	4.5	12.5	12.5	12.5	12.5	52.5	52.5				
62	2.5	112.5	14.5	30.5	72.5	72.5	14.5	30.5	72.5	72.5	42.5	46.5		s1-1-111/4		
63	2.5	124.5	6.5	6.5	6.5	6.5	2.5	2.5	2.5	2.5	26.5	26.5				
64	4.5	124.5	4.5	4.5	4.5	4.5	124.5	124.5	124.5	124.5	12.5	12.5				
65	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
66	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
67	16.5	127.5	87.5	87.5	87.5	87.5	110.5	110.5	110.5	110.5	56.5	56.5		s00-----/32		
68	4.5	116.5	4.5	4.5	4.5	4.5	60.5	60.5	60.5	60.5	4.5	4.5				
69	2.5	126.5	6.5	6.5	6.5	6.5	2.5	2.5	2.5	2.5	16.5	16.5				
70	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
71	4.5	124.5	22.5	22.5	22.5	22.5	18.5	18.5	18.5	18.5	30.5	30.5		s10--11-/8		
72	4.5	120.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	56.5	56.5				
73	2.5	112.5	35.5	35.5	35.5	35.5	35.5	35.5	35.5	35.5	88.5	88.5		s01-110-/4		
74	6.5	127.5	49.5	49.5	6.5	6.5	78.0	78.0	14.5	14.5	21.5	21.5				
75	4.5	126.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	24.5	24.5				
76	4.5	124.5	36.5	36.5	36.5	36.5	28.5	28.5	28.5	28.5	48.5	48.5		s01--11-/8		
77	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
78	2.0	120.5	6.0	6.0	13.5	13.5	2.0	2.0	15.5	15.5	57.5	57.5				
79	16.5	126.5	74.5	74.5	74.5	74.5	78.5	78.5	78.5	78.5	38.5	38.5		s00-----/32		
80	4.5	122.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	46.5	46.5		s11--11-/8		
81	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
82	4.0	120.5	20.5	20.5	90.5	90.5	4.0	4.0	68.5	68.5	95.5	95.5				
83	32.5	120.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	88.5	88.5		s0------/64		
84	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
85	16.5	124.5	54.5	54.5	16.5	16.5	110.5	110.5	16.5	16.5	34.5	34.5				
86	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
87	4.5	112.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	22.5	22.5				
88	1.5	112.5	66.5	66.5	66.5	66.5	82.5	82.5	82.5	82.5	82.5	82.5		s111110-/2		
89	2.0	120.5	83.5	83.5	109.5	109.5	2.0	2.0	14.5	14.5	105.5	105.5				
90	8.5	88.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	88.5	88.5		s1---11-/16		
91	16.5	127.5	38.5	38.5	38.5	38.5	62.5	62.5	62.5	62.5	34.5	34.5		s00-----/32		
92	32.5	112.5	112.5	112.5	32.5	32.5	112.5	112.5	32.5	32.5	80.5	80.5				
93	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
94	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
95	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
96	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
97	1.5	126.5	19.5	19.5	9.5	9.5	3.5	3.5	1.5	1.5	70.5	70.5				
98	4.5	112.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5		s-1-110-/8		
99	2.5	120.5	8.5	8.5	60.5	60.5	2.5	2.5	14.5	14.5	68.5	68.5				
AVE	-	-	46.2	46.4	48.9	48.9	50.6	50.8	52.1	52.1	55.0	55.0			-	-

KEY:
1: s1010101 2: s1000101 3: s1010100 4: s1000100 5: s1011101 6: s1001101 7: s1011100
8: s1001100 9: s1010001 10: s1000001 C:

Configurations

Numeric configuration analysis

CONFIGURATIONS																	
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
1	s1010101	23.4	100.0	76.6	82.2	0	0	1	0	5	9	13	9	19	14	30	18.60
2	s1000101	23.4	100.0	76.6	82.1	0	0	1	0	5	9	13	9	19	14	30	18.66
3	s1011101	20.0	100.0	80.0	81.7	0	1	0	0	4	11	12	12	19	12	29	18.30
4	s1001101	20.0	100.0	80.0	81.6	0	1	0	0	4	11	12	12	19	12	29	18.37
5	s1010100	23.4	100.0	76.6	80.7	0	0	2	1	6	9	15	9	11	14	33	20.76
6	s1000100	23.4	100.0	76.6	80.7	0	0	2	1	6	9	15	9	11	14	33	20.76
7	s1011100	20.0	100.0	80.0	80.2	0	1	0	1	6	9	11	18	16	9	29	19.17
8	s1001100	20.0	100.0	80.0	80.2	0	1	0	1	6	9	11	18	16	9	29	19.17
9	s1010111	20.0	100.0	80.0	79.0	0	1	1	0	6	18	10	6	14	16	28	20.40
10	s1011111	20.0	100.0	80.0	78.9	0	1	0	1	6	18	10	6	14	16	28	20.42
11	s1000111	20.0	100.0	80.0	78.9	0	1	1	0	6	19	9	6	14	16	28	20.46
12	s1001111	20.0	100.0	80.0	78.9	0	1	0	1	6	19	9	6	14	16	28	20.47
13	s0001101	16.7	100.0	83.3	78.8	0	1	3	5	4	13	5	11	13	8	37	23.24
14	s0011101	16.7	100.0	83.3	78.8	0	1	3	5	4	13	5	11	13	8	37	23.24
15	s1100101	36.4	100.0	63.6	78.5	0	0	0	1	5	11	17	14	22	10	20	17.58
16	s1110101	29.9	100.0	70.1	78.5	0	0	1	0	5	12	17	12	23	10	20	17.89
17	s1010001	20.0	100.0	80.0	78.3	0	1	0	3	10	11	11	6	16	13	29	21.77
18	s1000001	20.0	100.0	80.0	78.3	0	1	0	3	10	11	11	6	16	13	29	21.78
19	s1010110	20.0	100.0	80.0	78.1	0	1	1	3	7	17	8	6	12	14	31	21.89
20	s1000110	20.0	100.0	80.0	78.1	0	1	1	3	7	17	8	6	12	14	31	21.90
21	s1011110	20.0	100.0	80.0	78.1	0	1	0	4	7	17	8	6	12	14	31	21.95
22	s1001110	20.0	100.0	80.0	78.1	0	1	0	4	7	17	8	6	12	14	31	21.95
23	s1010000	20.0	100.0	80.0	78.0	0	1	1	5	8	13	8	6	15	11	32	22.86
24	s1000000	20.0	100.0	80.0	78.0	0	1	1	5	8	13	8	6	15	11	32	22.86
25	s1011001	20.0	100.0	80.0	78.0	0	1	0	2	8	13	13	10	15	10	28	20.63
26	s1001001	20.0	100.0	80.0	78.0	0	1	0	2	8	13	13	10	15	10	28	20.63
27	s1100100	36.4	100.0	63.6	77.8	0	0	0	1	6	7	24	14	18	10	20	17.75
28	s0010101	16.7	100.0	83.3	77.8	0	1	3	5	5	12	8	10	13	7	36	23.41
29	s0000101	16.7	100.0	83.3	77.8	0	1	3	5	5	12	8	10	13	7	36	23.41
30	s1110100	29.9	100.0	70.1	77.7	0	0	1	0	6	8	25	11	19	10	20	18.13
31	s0111101	28.6	100.0	71.4	77.6	0	0	1	0	9	15	12	13	15	10	25	19.31
32	s1011000	20.0	100.0	80.0	77.6	0	1	0	6	6	13	10	9	16	9	30	22.00
33	s1001000	20.0	100.0	80.0	77.6	0	1	0	6	6	13	10	9	16	9	30	22.00
34	s1110001	37.5	100.0	62.5	77.6	0	0	0	1	8	13	15	10	20	16	17	18.49
35	s1100001	37.0	100.0	63.0	77.5	0	0	0	1	8	13	15	9	21	16	17	18.56
36	s0000111	16.7	100.0	83.3	77.4	0	1	3	4	5	14	10	10	9	8	36	23.38
37	s0010111	16.7	100.0	83.3	77.4	0	1	3	4	5	14	10	10	9	8	36	23.38
38	s0011111	16.7	100.0	83.3	77.3	0	1	3	4	5	14	10	10	9	8	36	23.39
39	s0001111	16.7	100.0	83.3	77.3	0	1	3	4	5	14	10	10	9	8	36	23.39
40	s0110101	28.6	100.0	71.4	77.3	0	0	1	0	7	18	13	9	20	9	23	19.22
41	s0101101	28.6	100.0	71.4	77.2	0	0	1	1	9	13	13	14	14	11	24	19.59
42	s1111101	25.7	100.0	74.3	77.1	0	0	1	2	5	14	13	18	15	12	20	19.28
43	s1100000	34.8	100.0	65.2	77.0	0	0	0	3	7	13	15	8	21	15	18	19.38
44	s1110000	34.8	100.0	65.2	77.0	0	0	0	3	7	13	14	10	20	15	18	19.32
45	s0100101	28.6	100.0	71.4	77.0	0	0	1	1	7	16	14	9	20	9	23	19.62
46	s1101101	37.8	100.0	62.2	76.9	0	0	0	1	7	13	14	20	15	11	19	18.36
47	s1101100	37.8	100.0	62.2	76.5	0	0	0	1	7	10	18	24	13	8	19	17.86
48	s0110001	27.1	100.0	72.9	76.4	0	0	2	5	7	12	12	9	17	16	20	21.21
49	s0100001	28.6	100.0	71.4	76.3	0	0	2	5	7	12	12	9	17	16	20	21.19
50	s1111100	25.7	100.0	74.3	76.3	0	0	1	2	5	11	17	23	14	7	20	18.69
continued below																	

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile																		
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
51	s0111100	28.6	100.0	71.4	76.3	0	0	1	3	8	12	15	16	12	8	25	19.98
52	s0011001	16.7	100.0	83.3	76.2	0	1	3	6	9	11	9	9	8	8	36	24.45
53	s0001001	16.7	100.0	83.3	76.2	0	1	3	6	9	11	9	9	8	8	36	24.45
54	s0110100	28.6	100.0	71.4	76.1	0	0	1	4	7	12	18	10	16	9	23	20.27
55	s1111001	40.7	100.0	59.3	76.0	0	0	0	0	6	20	14	13	20	9	18	18.13
56	s0000001	16.7	100.0	83.3	76.0	0	1	3	8	9	8	10	8	8	9	36	25.17
57	s0010001	16.7	100.0	83.3	76.0	0	1	3	8	9	8	10	8	8	9	36	25.17
58	s0101100	28.6	100.0	71.4	75.9	0	0	1	4	8	10	16	16	12	9	24	20.25
59	s1101001	40.7	100.0	59.3	75.9	0	0	0	0	6	22	12	13	19	11	17	18.15
60	s0100100	28.6	100.0	71.4	75.8	0	0	1	5	8	9	19	10	16	9	23	20.60
61	s0110000	27.1	100.0	72.9	75.8	0	0	3	6	7	10	13	8	17	15	21	22.07
62	s0100000	28.0	100.0	72.0	75.8	0	0	3	6	7	10	13	8	17	15	21	22.05
63	s0010011	16.7	100.0	83.3	75.8	0	1	3	6	11	9	9	9	9	7	36	24.91
64	s0000011	16.7	100.0	83.3	75.8	0	1	3	6	11	9	9	9	9	7	36	24.91
65	s0111001	27.1	100.0	72.9	75.7	0	0	2	3	8	12	14	12	18	11	20	20.33
66	s0001011	16.7	100.0	83.3	75.7	0	1	3	6	11	9	9	9	9	7	36	24.93
67	s0011011	16.7	100.0	83.3	75.7	0	1	3	6	11	9	9	9	9	7	36	24.93
68	s0101001	28.6	100.0	71.4	75.7	0	0	2	3	8	13	13	12	18	11	20	20.32
69	s0110111	28.6	100.0	71.4	75.6	0	0	1	0	9	20	11	15	13	8	23	19.46
70	s0111111	28.6	100.0	71.4	75.6	0	0	1	0	9	20	11	15	13	8	23	19.47
71	s1101000	34.8	100.0	65.2	75.5	0	0	0	1	7	19	13	12	21	9	18	18.85
72	s0011000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
73	s0011010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
74	s0010110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
75	s0010100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
76	s0010010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
77	s0011110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
78	s0011100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
79	s0010000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
80	s0001110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
81	s0001100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
82	s0000110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
83	s0000100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
84	s0001010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
85	s0001000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
86	s0000000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
87	s0000010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
88	s1111000	34.8	100.0	65.2	75.5	0	0	0	1	7	19	13	12	22	8	18	18.82
89	s0100111	28.6	100.0	71.4	75.4	0	0	1	0	10	19	11	15	13	8	23	19.67
90	s0101111	28.6	100.0	71.4	75.4	0	0	1	0	10	19	11	15	13	8	23	19.68
91	s1100111	40.8	100.0	59.2	75.4	0	0	0	0	10	23	10	9	18	10	20	19.06
92	s1101111	40.8	100.0	59.2	75.4	0	0	0	0	10	23	10	9	18	10	20	19.05
93	s1010010	20.0	100.0	80.0	75.3	0	1	0	5	6	24	8	4	13	8	31	23.01
94	s1000010	20.0	100.0	80.0	75.3	0	1	0	5	6	24	8	4	13	8	31	23.01
95	s1110111	26.9	100.0	73.1	75.3	0	0	1	0	10	17	17	7	16	12	20	19.55
96	s1111111	28.5	100.0	71.5	75.3	0	0	1	0	10	17	17	7	16	12	20	19.53
97	s1011010	20.0	100.0	80.0	75.2	0	1	0	5	7	23	8	4	13	8	31	23.05
98	s1001010	20.0	100.0	80.0	75.2	0	1	0	5	7	23	8	4	13	8	31	23.05
99	s0111000	27.1	100.0	72.9	75.1	0	0	3	4	9	9	15	11	18	10	21	21.25
100	s0101000	28.0	100.0	72.0	75.1	0	0	3	4	9	9	15	11	18	10	21	21.24
continued below																	

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
101	s1010011	20.0	100.0	80.0	75.1	0	1	0	3	7	24	9	5	15	8	28	22.08
102	s1000011	20.0	100.0	80.0	75.1	0	1	0	3	7	24	9	5	15	8	28	22.09
103	s1011011	20.0	100.0	80.0	75.0	0	1	0	3	8	23	9	5	15	8	28	22.14
104	s1001011	20.0	100.0	80.0	75.0	0	1	0	3	8	23	9	5	15	8	28	22.15
105	s1101110	38.5	100.0	61.5	74.9	0	0	0	1	11	18	11	14	16	9	20	19.11
106	s1100110	38.5	100.0	61.5	74.9	0	0	0	1	11	18	11	14	16	9	20	19.13
107	s1110110	26.9	100.0	73.1	74.7	0	0	1	1	10	16	15	11	15	11	20	19.64
108	s1111110	28.5	100.0	71.5	74.7	0	0	1	1	10	16	15	11	15	11	20	19.62
109	s0111110	28.6	100.0	71.4	74.7	0	0	1	4	8	15	15	14	12	8	23	20.28
110	s0110110	28.6	100.0	71.4	74.7	0	0	1	4	8	15	15	14	12	8	23	20.28
111	s0101110	28.6	100.0	71.4	74.5	0	0	1	4	9	14	15	14	12	8	23	20.47
112	s0100110	28.6	100.0	71.4	74.5	0	0	1	4	9	14	15	14	12	8	23	20.47
113	s0110011	28.6	100.0	71.4	74.1	0	0	1	4	11	14	12	13	14	11	20	20.88
114	s0111011	28.6	100.0	71.4	74.1	0	0	1	4	11	14	12	13	14	11	20	20.90
115	s0100011	28.6	100.0	71.4	74.1	0	0	1	4	11	14	12	13	14	11	20	20.87
116	s0101011	28.6	100.0	71.4	74.1	0	0	1	4	11	14	12	13	14	11	20	20.89
117	s0111010	28.0	100.0	72.0	73.9	0	0	2	6	9	12	14	11	14	11	21	21.68
118	s0110010	28.0	100.0	72.0	73.9	0	0	2	6	9	12	14	11	14	11	21	21.68
119	s0101010	28.0	100.0	72.0	73.8	0	0	2	6	9	12	14	11	14	11	21	21.67
120	s0100010	28.0	100.0	72.0	73.8	0	0	2	6	9	12	14	11	14	11	21	21.67
121	s1100011	39.4	100.0	60.6	73.3	0	0	0	2	7	26	11	10	19	8	17	19.18
122	s1110011	39.6	100.0	60.4	73.3	0	0	0	1	8	25	13	10	18	8	17	19.12
123	s1100010	34.8	100.0	65.2	73.3	0	0	0	3	8	23	11	11	18	8	18	19.78
124	s1110010	34.8	100.0	65.2	73.2	0	0	0	2	9	23	12	10	18	8	18	19.73
125	s1101011	39.4	100.0	60.6	73.2	0	0	0	2	7	26	11	10	19	8	17	19.24
126	s1101010	34.8	100.0	65.2	73.2	0	0	0	3	8	23	11	11	18	8	18	19.82
127	s1111011	38.9	100.0	61.1	73.2	0	0	0	2	7	25	13	10	18	8	17	19.18
128	s1111010	34.8	100.0	65.2	73.1	0	0	0	3	8	23	12	10	18	8	18	19.77
	AVE	25.1	100.0	74.9	76.3	0.0	0.5	1.5	3.8	7.3	14.2	11.6	10.3	14.1	9.8	26.9	21.47

Ranked configuration analysis

CONFIGURATIONS
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
1	s1010101	2	124	122	46.2	41	15	17	9	6	4	2	0	1	0	7	32.76
2	s1000101	2	124	122	46.4	41	15	16	9	5	6	2	0	1	0	7	32.64
3	s1010100	2	124	122	48.9	45	14	11	8	5	5	1	1	1	0	12	34.02
4	s1000100	2	124	122	48.9	45	14	11	8	4	6	1	1	1	0	12	34.00
5	s1011101	2	124	122	50.6	48	15	8	10	3	1	5	2	2	1	5	38.10
6	s1001101	2	124	122	50.8	47	16	7	10	3	2	5	2	2	1	5	37.90
7	s1011100	1	124	123	52.1	46	15	5	9	5	1	6	3	2	0	9	36.67
8	s1001100	1	124	123	52.1	45	16	5	9	5	1	6	3	2	0	9	36.61
9	s1010001	4	116	112	55.0	34	18	12	9	3	2	5	1	2	0	14	26.02
10	s1000001	4	116	112	55.0	34	18	12	9	3	2	5	1	1	1	14	26.03
11	s1010111	2	124	122	55.7	40	11	13	10	6	3	4	1	1	1	14	36.66
12	s1011111	2	124	122	55.9	40	11	13	10	6	3	3	1	2	1	14	36.77
13	s1000111	4	124	120	56.0	38	11	14	11	6	3	4	1	1	1	14	36.41
14	s1001111	4	124	120	56.2	38	11	14	11	6	3	3	1	2	1	14	36.52
15	s1000000	4	127	123	56.3	38	15	11	8	4	2	3	2	1	0	18	29.02
16	s1010000	4	127	123	56.3	38	15	11	8	4	2	3	2	1	0	18	29.02
17	s1010110	3	125	122	57.3	42	11	8	9	6	3	4	3	1	1	14	35.12
18	s1011110	3	125	122	57.4	42	11	8	9	6	3	3	4	1	1	14	35.18
19	s1000110	4	127	123	57.4	41	11	9	9	6	3	4	3	1	1	14	35.06
20	s1001110	4	127	123	57.5	41	11	9	9	6	3	3	4	1	1	14	35.12
21	s1011001	8	124	116	57.5	33	12	16	6	5	6	3	2	3	0	15	26.58
22	s1001001	8	124	116	57.6	33	12	16	6	5	6	3	2	2	1	15	26.57
23	s1110101	4	125	121	57.7	29	17	17	13	4	3	2	1	3	3	10	40.18
24	s1110100	4	125	121	58.5	30	16	13	12	4	4	2	2	5	1	14	38.11
25	s1100101	4	124	120	58.7	29	18	11	13	4	7	3	0	3	4	9	39.04
26	s1011000	8	124	116	59.1	36	11	14	6	6	5	1	4	1	0	19	28.15
27	s1001000	8	124	116	59.1	36	11	14	6	6	5	1	4	1	0	19	28.15
28	s1100100	4	124	120	59.2	30	17	9	12	4	6	3	1	6	2	12	37.15
29	s0011101	2	126	124	60.0	45	6	7	12	4	3	5	2	5	1	10	32.55
30	s0001101	2	126	124	60.0	45	6	7	12	4	3	5	2	5	1	10	32.55
31	s1100001	8	116	108	60.8	22	25	7	11	5	8	4	3	1	1	14	22.53
32	s1110001	8	120	112	61.0	22	25	10	7	6	6	4	5	1	2	13	24.07
33	s0010101	12	126	114	62.1	43	6	6	12	4	3	4	2	5	3	12	30.05
34	s0000101	12	126	114	62.1	43	6	6	12	4	3	4	2	5	3	12	30.05
35	s1111100	1	127	126	62.1	33	14	12	9	8	2	2	6	1	2	11	41.50
36	s1100000	8	124	116	62.4	24	23	9	8	5	8	3	4	1	0	17	23.91
37	s1111101	1	127	126	62.5	37	13	14	9	6	1	1	4	2	1	12	44.09
38	s0010111	8	120	112	62.7	43	6	7	10	3	3	4	3	5	2	15	30.53
39	s0000111	8	120	112	62.7	43	6	7	10	3	3	4	3	5	2	15	30.53
40	s1110000	8	124	116	62.8	24	23	11	5	6	7	3	5	0	2	16	25.07
41	s1101100	4	127	123	62.8	31	15	12	9	8	1	2	7	0	3	14	40.90
42	s0001111	8	124	116	62.8	43	6	7	10	3	3	4	2	6	2	15	30.68
43	s0011111	8	124	116	62.8	43	6	7	10	3	3	4	2	6	2	15	30.68
44	s1101101	2	127	125	63.4	33	15	13	10	5	2	2	4	0	3	15	43.26
45	s0011001	16	122	106	64.6	43	7	5	10	1	3	4	2	5	1	20	29.20
46	s0001001	16	122	106	64.6	43	7	5	10	1	3	4	2	5	1	20	29.20
47	s0110101	1	124	123	65.0	32	16	9	11	5	2	2	6	1	5	12	34.34
48	s1010010	6	121	115	65.0	35	6	14	5	11	2	1	5	2	3	17	29.93
49	s1010011	12	120	108	65.1	31	7	16	5	10	4	3	5	1	2	16	28.07
50	s1000010	6	123	117	65.1	35	6	14	5	11	2	1	4	3	3	17	29.90

continued below

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
51	s1000011	12	120	108	65.2	31	7	16	5	10	4	2	5	2	2	16	28.00
52	s0010001	16	122	106	65.3	43	8	4	11	1	3	4	3	5	1	17	29.42
53	s0000001	16	122	106	65.3	43	8	4	11	1	3	4	3	5	1	17	29.42
54	s0100101	1	124	123	65.5	32	17	7	12	5	2	1	7	1	5	12	34.09
55	s1011010	6	124	118	65.6	35	6	14	4	11	3	1	4	3	3	17	30.50
56	s1101001	8	126	118	65.6	22	17	15	9	2	7	4	2	3	3	19	28.20
57	s0111101	2	127	125	65.7	36	12	11	9	7	3	3	3	0	4	14	36.95
58	s1001010	6	124	118	65.7	35	6	14	4	11	3	1	3	4	3	17	30.47
59	s1011011	12	124	112	65.7	31	7	16	4	10	4	4	4	2	2	16	28.80
60	s1001011	12	124	112	65.8	31	7	16	4	10	4	3	4	3	2	16	28.73
61	s0010011	16	112	96	65.8	43	6	5	10	1	5	4	3	5	1	17	28.59
62	s0000011	16	112	96	65.8	43	6	5	10	1	5	4	3	5	1	17	28.59
63	s1111001	4	126	122	65.9	23	16	15	8	3	6	4	4	4	4	15	29.72
64	s0011011	16	112	96	65.9	43	6	5	10	1	4	5	3	5	1	17	28.70
65	s0001011	16	112	96	65.9	43	6	5	10	1	4	5	3	5	1	17	28.70
66	s0101101	2	127	125	66.0	35	13	9	10	8	3	3	3	0	4	14	36.79
67	s0011010	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
68	s0011000	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
69	s0011100	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
70	s0010000	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
71	s0010110	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
72	s0010100	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
73	s0010010	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
74	s0001100	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
75	s0011110	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
76	s0001010	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
77	s0001000	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
78	s0000110	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
79	s0001110	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
80	s0000000	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
81	s0000100	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
82	s0000010	16	120	104	66.4	46	5	4	11	2	3	3	3	5	1	17	32.42
83	s1110111	2	126	124	67.0	30	10	14	8	10	5	2	3	3	1	17	41.88
84	s0110100	4	124	120	67.1	32	15	8	10	5	2	2	8	2	3	14	32.54
85	s1110110	4	126	122	67.1	29	10	12	8	11	6	3	2	3	1	17	38.05
86	s1111111	2	126	124	67.1	30	10	14	8	10	5	1	3	4	1	17	41.98
87	s1111110	4	126	122	67.1	29	10	12	8	11	6	2	3	3	1	17	38.04
88	s1101000	8	126	118	67.2	24	14	14	9	4	6	3	3	3	3	21	26.12
89	s0100001	4	122	118	67.3	27	21	4	11	6	8	3	0	2	1	19	20.86
90	s0100100	4	124	120	67.5	32	16	6	11	5	2	1	9	2	3	14	32.35
91	s1111000	8	126	118	67.5	24	14	15	7	5	5	3	5	3	4	18	27.56
92	s0110111	4	124	120	67.6	31	11	9	11	6	6	5	1	0	4	20	32.89
93	s0111111	4	124	120	67.7	31	11	9	11	6	6	5	0	1	4	20	33.02
94	s0110001	4	122	118	67.7	27	21	4	11	8	5	3	1	2	1	19	21.67
95	s0111100	2	124	122	67.9	36	10	9	7	6	4	5	4	0	4	18	34.01
96	s1100110	4	127	123	67.9	28	9	12	6	12	6	6	2	2	3	17	36.17
97	s1101110	4	124	120	68.0	28	9	12	6	12	6	5	3	2	3	17	36.20
98	s0101001	20	123	103	68.0	26	14	8	14	3	8	7	1	3	1	16	23.07
99	s1100111	4	127	123	68.1	28	9	13	9	10	5	6	2	2	2	18	39.75
100	s0101100	2	124	122	68.2	35	11	8	8	6	3	5	4	1	4	18	33.80

continued below

CONFIGURATIONS (continued)																	
/hame/jeremyc/phd/eval/sys/scores/DEV1/scorefile.ranked																	
NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
101	s0100111	4	124	120	68.2	30	12	9	11	5	7	5	1	0	4	18	32.41
102	s1101111	4	124	120	68.2	28	9	13	9	10	5	5	2	3	2	18	39.82
103	s0101111	4	124	120	68.3	30	12	9	11	5	7	5	0	1	4	18	32.54
104	s0111001	20	123	103	68.4	26	14	8	14	5	5	7	2	3	1	16	23.95
105	s0100000	4	122	118	68.5	29	19	5	9	6	8	3	1	1	2	18	22.18
106	s0110000	4	122	118	68.9	29	19	5	9	7	6	3	2	2	1	18	22.79
107	s0111110	4	124	120	69.7	32	10	8	10	4	7	4	2	1	4	23	29.50
108	s0110110	4	124	120	69.7	32	10	8	10	4	7	4	2	1	4	23	29.50
109	s0101000	26	122	96	69.7	28	12	8	13	5	7	6	1	1	2	19	22.35
110	s0111000	26	122	96	70.1	28	12	8	13	6	5	6	2	2	1	19	23.00
111	s0101110	4	124	120	70.2	31	11	8	10	4	7	4	2	0	5	21	29.16
112	s0100110	4	124	120	70.2	31	11	8	10	4	7	4	2	0	5	21	29.16
113	s0100011	32	120	88	72.0	26	13	7	13	6	9	2	2	3	1	20	21.01
114	s0101011	32	120	88	72.1	26	13	7	13	6	8	3	2	3	1	20	21.13
115	s0110011	32	120	88	72.3	26	13	7	13	8	6	2	3	3	1	20	21.75
116	s0111011	32	120	88	72.4	26	13	7	13	8	5	3	3	3	1	20	21.87
117	s0100010	32	120	88	72.6	28	12	7	12	7	8	2	1	1	2	23	22.12
118	s0101010	32	120	88	72.6	28	12	7	12	7	8	2	1	1	2	23	22.12
119	s0111010	32	120	88	72.9	28	12	7	12	8	6	2	2	2	1	23	22.67
120	s0110010	32	120	88	72.9	28	12	7	12	8	6	2	2	2	1	23	22.67
121	s1110010	16	124	108	74.0	22	10	13	7	13	5	3	6	2	3	18	26.69
122	s1100010	16	124	108	74.0	22	10	12	7	14	5	4	5	2	1	21	25.01
123	s1110011	16	124	108	74.2	20	11	14	7	11	5	4	8	2	4	15	27.11
124	s1100011	16	124	108	74.2	20	11	12	9	11	5	4	8	3	1	18	25.30
125	s1111010	16	124	108	74.4	22	10	13	6	13	6	3	5	3	3	18	27.17
126	s1101010	16	124	108	74.4	22	10	12	6	14	6	4	4	3	1	21	25.51
127	s1111011	16	124	108	74.6	20	11	14	6	11	5	5	7	3	4	15	27.71
128	s1101011	16	124	108	74.7	20	11	12	8	11	5	5	7	4	1	18	25.91
	AVE	10.3	123.3	113.0	64.5	34	11	9	9	5	4	3	2	2	1	16	30.89

Components

Numeric (left) and ranked (right) component analysis

sT------	1.111358	sT------	5.199063
s-L-----	-1.733543	s-L-----	-6.810312
s--S----	0.026074	s--S----	0.140469
s---P---	-0.247831	s---P---	-1.057969
s----C--	1.660815	s----C--	4.437187
s-----F-	-1.900357	s-----F-	-5.032813
s------U	0.804530	s------U	1.460938

C.3 Development corpus results II

This section contains the results tables for the development corpus using a LAM grammar modified following rule evaluation. The tables are presented in the same order as in the previous section.

Documents

Numeric document analysis

DOCUMENTS /hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
00	57.9	96.7	64.8	64.7	82.5	80.6	62.1	62.1	79.3	78.7	60.7	60.7		s01-0101/2	60.0	44.2
01	53.7	79.2	58.4	58.4	58.4	58.4	58.4	58.4	58.4	58.4	57.6	57.6		s11--10-/8	53.7	47.6
02	38.9	92.1	66.8	66.8	66.8	66.8	43.7	43.7	47.9	47.9	92.1	92.1			41.1	54.7
03	66.8	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	66.8	66.8			78.1	28.9
04	48.4	90.2	56.2	56.2	77.8	77.8	56.2	56.2	77.8	77.8	90.2	90.2			69.3	43.1
05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	19.3
06	46.2	99.1	99.1	99.1	91.7	91.7	99.1	99.1	91.7	91.7	64.6	64.6			72.9	32.9
07	47.1	100.0	88.2	88.2	88.2	88.2	88.2	88.2	88.2	88.2	47.1	47.1		s00-----/32	100.0	18.4
08	85.0	100.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	88.5	88.5		s1---0--/32	96.0	14.4
09	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	26.1
10	73.5	100.0	73.5	73.5	77.0	77.0	73.5	73.5	77.0	77.0	80.4	80.4		s0------/64	100.0	14.7
11	25.7	100.0	54.9	54.9	59.2	59.2	54.9	54.9	59.2	59.2	29.2	30.1		s00-----/32	100.0	10.2
12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	46.4
13	51.0	85.4	85.4	85.4	70.3	70.3	85.4	85.4	70.3	70.3	56.9	58.5			82.6	24.8
14	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0			87.0	24.1
15	89.5	100.0	94.9	94.9	97.4	97.4	94.9	94.9	97.4	97.4	94.7	94.7		s01--1--/16	89.5	19.4
16	55.1	100.0	89.7	89.7	55.1	55.1	89.7	89.7	55.1	55.1	58.3	56.6		s0------/64	100.0	8.2
17	42.6	100.0	96.4	96.4	86.3	86.3	96.4	96.4	86.3	86.3	100.0	100.0			96.4	13.0
18	50.0	64.0	60.3	60.3	64.0	64.0	60.3	60.3	64.0	64.0	57.4	57.4			51.5	48.5
19	85.4	100.0	96.9	96.9	85.4	85.4	96.9	96.9	85.4	85.4	85.4	85.4		s0------/64	100.0	12.2
20	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	10.1
21	70.8	100.0	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2		s00-----/32	100.0	15.8
22	41.2	90.3	41.2	41.2	41.2	41.2	41.2	41.2	41.2	41.2	82.6	82.6		s0------/64	90.3	18.7
23	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	17.9
24	36.9	68.0	62.6	62.6	68.0	68.0	62.6	62.6	68.0	68.0	58.6	58.6			36.9	62.3
25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	52.7
26	47.5	67.1	63.8	63.6	67.1	66.9	63.8	63.6	67.1	66.9	67.1	67.1			54.5	49.5
27	47.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	16.5
28	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	27.3
29	37.5	88.0	78.6	78.6	47.4	47.4	78.6	78.6	47.4	47.4	83.2	83.2		s01-0-0-/8	75.2	35.4
30	73.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.2
31	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	36.4
32	34.8	69.3	59.5	59.5	69.3	69.3	59.5	59.5	69.3	69.3	60.4	60.4			34.8	62.1
33	37.6	92.4	70.0	70.0	84.3	84.3	70.0	70.0	84.3	84.3	56.7	56.7		s01-110-/4	37.6	60.0
34	76.1	83.0	79.7	79.7	79.7	79.7	79.7	79.7	79.7	79.7	83.0	83.0			78.2	28.7
35	45.0	62.6	45.0	45.0	62.6	62.6	45.0	45.0	62.6	62.6	62.6	62.6			54.4	44.4
36	56.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	25.5
37	53.3	100.0	100.0	100.0	53.3	53.3	100.0	100.0	53.3	53.3	90.5	90.5			74.3	36.7
38	28.6	73.7	61.6	61.6	73.7	73.7	61.6	61.6	73.7	73.7	47.8	47.8			28.6	67.9
39	47.3	66.4	60.2	60.2	66.4	66.4	60.2	60.2	66.4	66.4	51.5	51.5			47.3	50.6
40	20.0	98.2	87.1	87.1	87.1	87.1	87.1	87.1	87.1	87.1	98.2	98.2			20.0	76.6
41	54.1	100.0	81.4	81.4	100.0	100.0	54.1	54.1	54.1	54.1	91.9	91.9			54.1	47.9
42	36.4	94.1	94.1	94.1	72.1	72.1	94.1	94.1	72.1	72.1	91.1	91.1			90.9	16.7
43	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6			95.6	14.8
44	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	7.1
45	72.9	93.7	88.0	88.0	88.0	88.0	88.0	88.0	88.0	88.0	93.7	93.7			72.9	36.8
46	48.8	55.6	53.9	53.9	55.6	55.6	53.9	53.9	55.6	55.6	52.4	52.4			48.8	56.1
47	48.2	84.1	65.6	65.6	77.2	77.2	65.6	65.6	77.2	77.2	48.2	48.2		s-------/72	84.1	27.5
48	35.9	100.0	82.1	82.1	82.1	82.1	35.9	35.9	35.9	35.9	82.1	82.1		s-1--1-1/16	35.9	53.8
49	89.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	19.9
continued below																

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
50	89.8	100.0	89.8	89.8	89.8	89.8	89.8	89.8	89.8	89.8	89.8	89.8		s-1--1--/32	89.8	25.9
51	84.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	23.1
52	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.7
53	50.8	88.1	84.9	84.9	88.1	88.1	61.8	61.8	61.8	61.8	51.8	50.8			61.8	40.4
54	53.8	97.3	97.3	97.3	84.7	84.7	97.3	97.3	84.7	84.7	54.8	53.8			97.3	16.6
55	67.8	100.0	82.0	82.0	82.0	82.0	100.0	100.0	100.0	100.0	67.8	67.8			100.0	13.8
56	32.3	78.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7	44.7		s-11-1--/16	32.3	66.5
57	52.9	75.7	64.3	64.3	64.3	64.3	64.3	64.3	64.3	64.3	55.2	55.2		s1---00-/16	60.0	41.0
58	48.8	100.0	91.1	91.1	83.1	83.1	91.1	91.1	83.1	83.1	95.5	95.5		s00-----/32	100.0	9.1
59	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	21.7
60	94.0	100.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0	94.0		s0------/64	100.0	10.7
61	69.0	100.0	100.0	100.0	99.0	99.0	100.0	100.0	99.0	99.0	93.6	93.6			69.0	33.5
62	37.5	64.1	59.5	54.8	59.5	54.8	42.5	42.5	42.5	42.5	64.1	64.1			37.5	61.3
63	51.8	98.4	92.9	92.9	98.4	98.4	92.9	92.9	98.4	98.4	54.2	54.2			62.8	39.1
64	51.5	88.9	88.9	88.9	51.5	51.5	88.9	88.9	51.5	51.5	81.6	81.6			80.5	24.5
65	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2			75.2	41.0
66	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	12.4
67	42.9	100.0	91.5	91.5	80.2	80.2	91.5	91.5	80.2	80.2	95.7	95.7		s00-----/32	100.0	10.0
68	53.3	100.0	100.0	100.0	81.9	81.9	100.0	100.0	81.9	81.9	74.3	74.3			90.5	22.4
69	41.6	73.2	66.4	66.4	73.2	73.2	66.4	66.4	73.2	73.2	49.3	49.3			64.1	41.9
70	53.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.1
71	49.4	94.7	74.5	74.5	78.0	78.0	74.5	74.5	78.0	78.0	94.7	94.7			58.5	42.2
72	46.2	84.6	84.6	84.6	84.6	84.6	84.6	84.6	84.6	84.6	46.2	46.2			53.8	51.3
73	25.3	70.6	54.9	54.9	54.9	54.9	54.9	54.9	54.9	54.9	51.3	51.3		s01-110-/4	25.3	60.7
74	44.9	100.0	73.5	73.5	64.7	64.7	100.0	100.0	89.0	89.0	73.5	73.5			77.2	32.4
75	54.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	93.9	93.9			83.0	27.7
76	46.2	84.6	64.0	64.0	70.2	70.2	64.0	64.0	70.2	70.2	80.9	80.9		s01--11-/8	56.0	45.5
77	89.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.0
78	24.6	92.9	83.1	83.1	92.9	92.9	71.1	71.1	70.8	70.8	60.0	60.0			24.6	67.7
79	40.5	100.0	82.8	82.8	82.6	82.6	82.8	82.8	82.6	82.6	86.3	86.3		s00-----/32	100.0	7.4
80	41.0	100.0	76.9	76.9	76.9	76.9	76.9	76.9	76.9	76.9	76.9	76.9		s11--11-/8	48.7	62.8
81	84.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	28.2
82	31.8	100.0	87.9	87.9	100.0	100.0	56.1	56.1	72.7	72.7	87.9	87.9			31.8	60.6
83	39.6	85.8	46.7	46.7	46.7	46.7	46.7	46.7	46.7	46.7	70.4	70.4		s0------/64	85.8	34.8
84	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.0
85	40.8	100.0	85.1	85.1	50.9	50.9	100.0	100.0	100.0	100.0	55.6	55.6			100.0	25.9
86	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.1
87	57.4	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	81.6	81.6			57.4	44.2
88	16.7	86.8	23.4	23.4	20.0	20.0	23.4	23.4	20.0	20.0	20.0	20.0		s111110-/2	16.7	75.5
89	24.2	98.8	42.5	42.5	98.8	98.8	28.1	28.1	71.0	71.0	54.8	53.0			24.2	74.3
90	47.3	100.0	63.7	63.7	63.7	63.7	63.7	63.7	63.7	63.7	100.0	100.0			47.3	49.5
91	46.1	74.6	70.1	70.1	65.0	65.0	70.1	70.1	65.0	65.0	53.9	55.2		s00-----/32	74.6	32.2
92	57.2	62.8	57.2	57.2	57.2	57.2	62.8	62.8	62.8	62.8	57.2	57.2			62.8	38.6
93	49.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	12.6
94	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	21.4
95	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.5
96	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	12.1
97	35.5	70.9	62.4	62.4	67.4	67.4	65.7	65.7	70.9	70.9	66.4	66.4			36.0	61.0
98	38.2	100.0	56.4	56.4	56.4	56.4	56.4	56.4	56.4	56.4	56.4	56.4		s-1-110-/8	38.2	49.1
99	31.2	92.8	82.2	82.2	92.8	92.8	65.2	65.2	78.6	78.6	76.8	76.8			31.2	62.0
AVE	60.5	92.1	82.2	82.1	81.7	81.6	80.7	80.7	80.2	80.2	79.0	78.9			75.5	33.5

KEY:
1: s1010101 2: s1000101 3: s1011101 4: s1001101 5: s1010100 6: s1000100 7: s1011100
8: s1001100 9: s1010111 10: s1011111 C:

Ranked document analysis

DOCUMENTS /hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile.ranked															
			OVERALL RANKED CONFIGURATIONS												
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[] [?]
00	1.5	127.5	77.0	78.0	93.0	93.0	29.0	36.0	39.0	40.0	91.0	93.0		s01-0101/2	
01	4.5	112.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5		s11--10-/8	
02	5.5	127.5	42.5	42.5	101.5	101.5	42.5	42.5	77.5	77.5	96.5	96.5		s1---1-1/10	
03	8.5	120.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	24.5	24.5			
04	8.5	124.5	96.5	96.5	96.5	96.5	20.5	20.5	20.5	20.5	96.5	96.5		s1---11-/16	
05	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
06	2.5	124.5	2.5	2.5	2.5	2.5	6.5	6.5	6.5	6.5	14.5	14.5			
07	16.5	124.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5		s00-----/32	
08	16.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	16.5	16.5			
09	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
10	32.5	124.5	124.5	124.5	124.5	124.5	116.5	116.5	116.5	116.5	80.5	80.5		s0------/64	
11	16.5	127.5	73.5	73.5	73.5	73.5	66.5	66.5	66.5	66.5	40.5	40.5		s00-----/32	
12	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
13	4.5	124.5	4.5	4.5	4.5	4.5	92.5	92.5	92.5	92.5	12.5	12.5			
14	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
15	8.5	112.5	26.5	26.5	26.5	26.5	18.5	18.5	18.5	18.5	88.5	88.5		s01--1--/16	
16	32.5	124.5	76.5	76.5	76.5	76.5	124.5	124.5	124.5	124.5	68.5	68.5		s0------/64	
17	12.5	127.5	42.5	42.5	42.5	42.5	92.5	92.5	92.5	92.5	12.5	12.5			
18	6.5	120.5	14.5	14.5	14.5	14.5	6.5	6.5	6.5	6.5	64.5	64.5			
19	32.5	116.5	72.5	72.5	72.5	72.5	116.5	116.5	116.5	116.5	72.5	72.5		s0------/64	
20	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
21	16.5	126.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5		s00-----/32	
22	32.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	96.5	96.5		s0------/64	
23	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
24	4.5	112.5	12.5	12.5	12.5	12.5	4.5	4.5	4.5	4.5	42.5	42.5			
25	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
26	3.5	124.5	17.5	19.5	17.5	19.5	3.5	7.5	3.5	7.5	50.5	50.5			
27	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
28	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
29	4.5	127.5	34.5	34.5	34.5	34.5	122.5	122.5	122.5	122.5	26.5	26.5		s01-0-0-/8	
30	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
31	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
32	2.5	112.5	22.5	22.5	22.5	22.5	2.5	2.5	2.5	2.5	70.5	70.5			
33	2.5	112.5	14.5	14.5	14.5	14.5	6.5	6.5	6.5	6.5	78.5	78.5		s01-110-/4	
34	8.5	120.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	24.5	24.5		s1---11-/16	
35	12.5	124.5	124.5	124.5	124.5	124.5	12.5	12.5	12.5	12.5	116.5	116.5			
36	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
37	4.5	124.5	4.5	4.5	4.5	4.5	124.5	124.5	124.5	124.5	40.5	40.5			
38	4.5	96.5	12.5	12.5	12.5	12.5	4.5	4.5	4.5	4.5	48.5	48.5			
39	4.5	96.5	20.5	20.5	20.5	20.5	4.5	4.5	4.5	4.5	28.5	28.5			
40	4.5	112.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	64.5	64.5		s10--11-/8	
41	2.5	112.5	26.5	26.5	112.5	112.5	2.5	2.5	112.5	112.5	94.5	94.5			
42	2.5	127.5	2.5	2.5	2.5	2.5	110.5	110.5	110.5	110.5	18.5	18.5			
43	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
44	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
45	8.5	96.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	48.5	48.5		s1---11-/16	
46	4.5	96.5	12.5	12.5	12.5	12.5	4.5	4.5	4.5	4.5	48.5	48.5			
47	36.5	120.5	92.5	92.5	92.5	92.5	80.5	80.5	80.5	80.5	80.5	80.5		s-------/72	
48	8.5	112.5	28.5	28.5	112.5	112.5	28.5	28.5	112.5	112.5	88.5	88.5		s-1--1-1/16	
49	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
continued below															

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile.ranked

			OVERALL RANKED CONFIGURATIONS												
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[] [?]
50	16.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5		s-1--1--/32	
51	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
52	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
53	4.5	124.5	12.5	12.5	64.5	64.5	4.5	4.5	64.5	64.5	28.5	28.5			
54	8.5	124.5	8.5	8.5	8.5	8.5	92.5	92.5	92.5	92.5	8.5	8.5			
55	32.5	120.5	104.5	104.5	32.5	32.5	104.5	104.5	32.5	32.5	72.5	72.5			
56	8.5	112.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	88.5	88.5		s-11-1--/16	
57	8.5	120.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	8.5	8.5			
58	16.5	124.5	58.5	58.5	58.5	58.5	62.5	62.5	62.5	62.5	44.5	44.5		s00-----/32	
59	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
60	32.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	80.5	80.5		s0------/64	
61	4.5	96.5	4.5	4.5	4.5	4.5	12.5	12.5	12.5	12.5	52.5	52.5			
62	2.5	112.5	14.5	30.5	72.5	72.5	14.5	30.5	72.5	72.5	42.5	46.5		s1-1-111/4	
63	2.5	124.5	6.5	6.5	6.5	6.5	2.5	2.5	2.5	2.5	26.5	26.5			
64	4.5	124.5	4.5	4.5	4.5	4.5	124.5	124.5	124.5	124.5	12.5	12.5			
65	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
66	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
67	16.5	127.5	87.5	87.5	87.5	87.5	110.5	110.5	110.5	110.5	56.5	56.5		s00-----/32	
68	4.5	116.5	4.5	4.5	4.5	4.5	60.5	60.5	60.5	60.5	4.5	4.5			
69	2.5	126.5	6.5	6.5	6.5	6.5	2.5	2.5	2.5	2.5	16.5	16.5			
70	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
71	4.5	124.5	22.5	22.5	22.5	22.5	18.5	18.5	18.5	18.5	30.5	30.5		s10--11-/8	
72	4.5	120.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	56.5	56.5			
73	2.5	112.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5	88.5	88.5		s01-110-/4	
74	6.5	127.5	49.5	49.5	6.5	6.5	78.0	78.0	14.5	14.5	21.5	21.5			
75	4.5	126.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	24.5	24.5			
76	4.5	124.5	36.5	36.5	36.5	36.5	28.5	28.5	28.5	28.5	48.5	48.5		s01--11-/8	
77	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
78	2.0	120.5	6.0	6.0	13.5	13.5	2.0	2.0	15.5	15.5	57.5	57.5			
79	16.5	126.5	74.5	74.5	74.5	74.5	78.5	78.5	78.5	78.5	38.5	38.5		s00-----/32	
80	4.5	122.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	46.5	46.5		s11--11-/8	
81	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
82	4.0	120.5	20.5	20.5	90.5	90.5	4.0	4.0	68.5	68.5	95.5	95.5			
83	32.5	120.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	88.5	88.5		s0------/64	
84	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
85	16.5	124.5	54.5	54.5	16.5	16.5	110.5	110.5	16.5	16.5	34.5	34.5			
86	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
87	8.5	96.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	44.5	44.5			
88	1.5	112.5	66.5	66.5	66.5	66.5	82.5	82.5	82.5	82.5	82.5	82.5		s111110-/2	
89	2.0	120.5	83.5	83.5	109.5	109.5	2.0	2.0	14.5	14.5	105.5	105.5			
90	8.5	88.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	88.5	88.5		s1---11-/16	
91	16.5	127.5	38.5	38.5	38.5	38.5	62.5	62.5	62.5	62.5	34.5	34.5		s00-----/32	
92	32.5	112.5	112.5	112.5	32.5	32.5	112.5	112.5	32.5	32.5	80.5	80.5			
93	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5			
94	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
95	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
96	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5			
97	1.5	126.5	19.5	19.5	9.5	9.5	3.5	3.5	1.5	1.5	70.5	70.5			
98	4.5	112.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5		s-1-110-/8	
99	2.5	112.5	6.5	6.5	54.5	54.5	2.5	2.5	10.5	10.5	62.5	62.5			
AVE	-	-	46.5	46.6	49.1	49.1	50.8	51.1	52.3	52.3	55.1	55.2			- -

KEY:
1: s1010101 2: s1000101 3: s1010100 4: s1000100 5: s1011101 6: s1001101 7: s1011100
8: s1001100 9: s1010001 10: s1000001 C:

Configurations

Numeric configuration analysis

CONFIGURATIONS
/hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile

NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
1	s1010101	23.4	100.0	76.6	82.2	0	0	1	0	5	9	13	9	19	14	30	18.60
2	s1000101	23.4	100.0	76.6	82.1	0	0	1	0	5	9	13	9	19	14	30	18.66
3	s1011101	20.0	100.0	80.0	81.7	0	1	0	0	4	11	12	12	19	12	29	18.30
4	s1001101	20.0	100.0	80.0	81.6	0	1	0	0	4	11	12	12	19	12	29	18.37
5	s1010100	23.4	100.0	76.6	80.7	0	0	2	1	6	9	15	9	11	14	33	20.76
6	s1000100	23.4	100.0	76.6	80.7	0	0	2	1	6	9	15	9	11	14	33	20.76
7	s1011100	20.0	100.0	80.0	80.2	0	1	0	1	6	9	11	18	16	9	29	19.17
8	s1001100	20.0	100.0	80.0	80.2	0	1	0	1	6	9	11	18	16	9	29	19.17
9	s1010111	20.0	100.0	80.0	79.0	0	1	1	0	6	18	10	6	14	16	28	20.40
10	s1011111	20.0	100.0	80.0	78.9	0	1	0	1	6	18	10	6	14	16	28	20.42
11	s1100101	36.4	100.0	63.6	78.9	0	0	0	1	5	11	17	13	21	11	21	17.83
12	s1110101	29.9	100.0	70.1	78.9	0	0	1	0	5	12	17	11	22	11	21	18.14
13	s1000111	20.0	100.0	80.0	78.9	0	1	1	0	6	19	9	6	14	16	28	20.46
14	s1001111	20.0	100.0	80.0	78.9	0	1	0	1	6	19	9	6	14	16	28	20.47
15	s0001101	16.7	100.0	83.3	78.8	0	1	3	5	4	13	5	11	13	8	37	23.24
16	s0011101	16.7	100.0	83.3	78.8	0	1	3	5	4	13	5	11	13	8	37	23.24
17	s1010001	20.0	100.0	80.0	78.3	0	1	0	3	10	11	11	6	16	13	29	21.77
18	s1000001	20.0	100.0	80.0	78.3	0	1	0	3	10	11	11	6	16	13	29	21.78
19	s1100100	36.4	100.0	63.6	78.2	0	0	0	1	6	7	24	13	17	11	21	18.01
20	s1010110	20.0	100.0	80.0	78.1	0	1	1	3	7	17	8	6	12	14	31	21.89
21	s1000110	20.0	100.0	80.0	78.1	0	1	1	3	7	17	8	6	12	14	31	21.90
22	s1011110	20.0	100.0	80.0	78.1	0	1	0	4	7	17	8	6	12	14	31	21.95
23	s1001110	20.0	100.0	80.0	78.1	0	1	0	4	7	17	8	6	12	14	31	21.95
24	s1110100	29.9	100.0	70.1	78.1	0	0	1	0	6	8	25	10	18	11	21	18.39
25	s1000000	20.0	100.0	80.0	78.0	0	1	1	5	8	13	8	6	15	11	32	22.86
26	s1010000	20.0	100.0	80.0	78.0	0	1	1	5	8	13	8	6	15	11	32	22.86
27	s1011001	20.0	100.0	80.0	78.0	0	1	0	2	8	13	13	10	15	10	28	20.63
28	s1001001	20.0	100.0	80.0	78.0	0	1	0	2	8	13	13	10	15	10	28	20.63
29	s0111101	28.6	100.0	71.4	77.8	0	0	1	0	10	14	10	15	15	9	26	19.40
30	s0000101	16.7	100.0	83.3	77.8	0	1	3	5	5	12	8	10	13	7	36	23.41
31	s0010101	16.7	100.0	83.3	77.8	0	1	3	5	5	12	8	10	13	7	36	23.41
32	s1110001	37.5	100.0	62.5	77.8	0	0	0	1	9	12	14	10	21	15	18	18.57
33	s1100001	37.0	100.0	63.0	77.7	0	0	0	1	9	12	14	9	22	15	18	18.64
34	s1111101	25.7	100.0	74.3	77.6	0	0	1	2	5	14	12	17	15	13	21	19.51
35	s1011000	20.0	100.0	80.0	77.6	0	1	0	6	6	13	10	9	16	9	30	22.00
36	s1001000	20.0	100.0	80.0	77.6	0	1	0	6	6	13	10	9	16	9	30	22.00
37	s1101101	37.8	100.0	62.2	77.5	0	0	0	1	7	13	13	19	15	12	20	18.60
38	s0101101	28.6	100.0	71.4	77.4	0	0	1	1	10	12	11	16	14	10	25	19.68
39	s0000111	16.7	100.0	83.3	77.4	0	1	3	4	5	14	10	10	9	8	36	23.38
40	s0010111	16.7	100.0	83.3	77.4	0	1	3	4	5	14	10	10	9	8	36	23.38
41	s0001111	16.7	100.0	83.3	77.3	0	1	3	4	5	14	10	10	9	8	36	23.39
42	s0011111	16.7	100.0	83.3	77.3	0	1	3	4	5	14	10	10	9	8	36	23.39
43	s0110101	28.6	100.0	71.4	77.3	0	0	1	0	8	17	14	9	18	9	24	19.39
44	s1100000	34.8	100.0	65.2	77.2	0	0	0	3	8	12	14	8	22	14	19	19.47
45	s1110000	34.8	100.0	65.2	77.2	0	0	0	3	8	12	13	10	21	14	19	19.41
46	s0100101	28.6	100.0	71.4	77.0	0	0	1	1	8	15	15	9	18	9	24	19.78
47	s1101100	37.8	100.0	62.2	77.0	0	0	0	1	7	10	17	23	13	9	20	18.12
48	s1111100	25.7	100.0	74.3	76.8	0	0	1	2	5	11	16	22	14	8	21	18.94
49	s0111100	28.6	100.0	71.4	76.4	0	0	1	4	8	11	13	17	12	8	26	20.42
50	s0110001	27.1	100.0	72.9	76.4	0	0	2	5	8	11	12	9	17	15	21	21.40

continued below

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile

NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
51	s0100001	28.6	100.0	71.4	76.3	0	0	2	5	8	11	12	9	17	15	21	21.38
52	s1111001	40.7	100.0	59.3	76.3	0	0	0	0	7	19	14	12	19	10	19	18.35
53	s0110100	28.6	100.0	71.4	76.2	0	0	1	4	8	11	18	10	15	9	24	20.56
54	s0011001	16.7	100.0	83.3	76.2	0	1	3	6	9	11	9	9	8	8	36	24.45
55	s0001001	16.7	100.0	83.3	76.2	0	1	3	6	9	11	9	9	8	8	36	24.45
56	s1101001	40.7	100.0	59.3	76.2	0	0	0	0	7	21	12	12	18	12	18	18.37
57	s0101100	28.6	100.0	71.4	76.1	0	0	1	5	8	9	14	17	12	9	25	20.69
58	s1100111	40.8	100.0	59.2	76.1	0	0	0	0	10	22	9	9	19	10	21	19.01
59	s1101111	40.8	100.0	59.2	76.1	0	0	0	0	10	22	9	9	19	10	21	18.99
60	s0110111	28.6	100.0	71.4	76.1	0	0	1	0	10	17	13	14	13	8	24	19.48
61	s0111111	28.6	100.0	71.4	76.0	0	0	1	0	10	17	13	14	13	8	24	19.49
62	s0000001	16.7	100.0	83.3	76.0	0	1	3	8	9	8	10	8	8	9	36	25.17
63	s0010001	16.7	100.0	83.3	76.0	0	1	3	8	9	8	10	8	8	9	36	25.17
64	s1110111	26.9	100.0	73.1	76.0	0	0	1	0	10	16	16	7	17	12	21	19.50
65	s1111111	28.5	100.0	71.5	76.0	0	0	1	0	10	16	16	7	17	12	21	19.48
66	s0100100	28.6	100.0	71.4	75.9	0	0	1	5	9	8	19	10	15	9	24	20.89
67	s0110000	27.1	100.0	72.9	75.8	0	0	3	6	7	10	13	8	17	14	22	22.25
68	s0100000	28.0	100.0	72.0	75.8	0	0	3	6	7	10	13	8	17	14	22	22.23
69	s0100111	28.6	100.0	71.4	75.8	0	0	1	0	11	16	13	14	13	8	24	19.69
70	s0101111	28.6	100.0	71.4	75.8	0	0	1	0	11	16	13	14	13	8	24	19.71
71	s1101000	34.8	100.0	65.2	75.8	0	0	0	1	8	18	13	11	20	10	19	19.07
72	s0111001	27.1	100.0	72.9	75.8	0	0	2	3	8	13	13	12	17	11	21	20.55
73	s0010011	16.7	100.0	83.3	75.8	0	1	3	6	11	9	9	9	9	7	36	24.91
74	s0000011	16.7	100.0	83.3	75.8	0	1	3	6	11	9	9	9	9	7	36	24.91
75	s1111000	34.8	100.0	65.2	75.8	0	0	0	1	8	18	13	11	21	9	19	19.04
76	s0001011	16.7	100.0	83.3	75.7	0	1	3	6	11	9	9	9	9	7	36	24.93
77	s0011011	16.7	100.0	83.3	75.7	0	1	3	6	11	9	9	9	9	7	36	24.93
78	s0101001	28.6	100.0	71.4	75.7	0	0	2	3	8	14	12	12	17	11	21	20.54
79	s1101110	38.5	100.0	61.5	75.6	0	0	0	1	11	17	10	14	17	9	21	19.07
80	s1100110	38.5	100.0	61.5	75.6	0	0	0	1	11	17	10	14	17	9	21	19.09
81	s0011100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
82	s0011110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
83	s0011010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
84	s0000000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
85	s0001110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
86	s0011000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
87	s0001100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
88	s0010100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
89	s0001010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
90	s0000010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
91	s0010000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
92	s0000100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
93	s0010010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
94	s0010110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
95	s0001000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
96	s0000110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
97	s1110110	26.9	100.0	73.1	75.4	0	0	1	1	10	15	14	11	16	11	21	19.61
98	s1111110	28.5	100.0	71.5	75.4	0	0	1	1	10	15	14	11	16	11	21	19.59
99	s1010010	20.0	100.0	80.0	75.3	0	1	0	5	6	24	8	4	13	8	31	23.01
100	s1000010	20.0	100.0	80.0	75.3	0	1	0	5	6	24	8	4	13	8	31	23.01

continued below

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile																		
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
101	s0111000	27.1	100.0	72.9	75.2	0	0	3	5	8	9	15	11	17	10	22	21.61
102	s1011010	20.0	100.0	80.0	75.2	0	1	0	5	7	23	8	4	13	8	31	23.05
103	s1001010	20.0	100.0	80.0	75.2	0	1	0	5	7	23	8	4	13	8	31	23.05
104	s0101000	28.0	100.0	72.0	75.2	0	0	3	5	8	9	15	11	17	10	22	21.60
105	s1010011	20.0	100.0	80.0	75.1	0	1	0	3	7	24	9	5	15	8	28	22.08
106	s1000011	20.0	100.0	80.0	75.1	0	1	0	3	7	24	9	5	15	8	28	22.09
107	s1011011	20.0	100.0	80.0	75.0	0	1	0	3	8	23	9	5	15	8	28	22.14
108	s0110110	28.6	100.0	71.4	75.0	0	0	1	5	9	11	16	14	12	8	24	20.63
109	s0111110	28.6	100.0	71.4	75.0	0	0	1	5	9	11	16	14	12	8	24	20.63
110	s1001011	20.0	100.0	80.0	75.0	0	1	0	3	8	23	9	5	15	8	28	22.15
111	s0101110	28.6	100.0	71.4	74.8	0	0	1	5	10	10	16	14	12	8	24	20.82
112	s0100110	28.6	100.0	71.4	74.8	0	0	1	5	10	10	16	14	12	8	24	20.82
113	s0110011	28.6	100.0	71.4	74.3	0	0	1	4	12	14	11	12	14	11	21	21.21
114	s0111011	28.6	100.0	71.4	74.3	0	0	1	4	12	14	11	12	14	11	21	21.23
115	s0100011	28.6	100.0	71.4	74.2	0	0	1	4	12	14	11	12	14	11	21	21.20
116	s0101011	28.6	100.0	71.4	74.2	0	0	1	4	12	14	11	12	14	11	21	21.22
117	s0111010	28.0	100.0	72.0	74.0	0	0	2	7	9	11	14	10	14	11	22	22.12
118	s0110010	28.0	100.0	72.0	74.0	0	0	2	7	9	11	14	10	14	11	22	22.12
119	s0100010	28.0	100.0	72.0	73.9	0	0	2	7	9	11	14	10	14	11	22	22.12
120	s0101010	28.0	100.0	72.0	73.9	0	0	2	7	9	11	14	10	14	11	22	22.12
121	s1100011	39.4	100.0	60.6	73.7	0	0	0	2	7	26	10	9	20	8	18	19.38
122	s1110011	39.6	100.0	60.4	73.7	0	0	0	1	8	25	12	9	19	8	18	19.32
123	s1100010	34.8	100.0	65.2	73.7	0	0	0	3	8	23	10	10	19	8	19	19.98
124	s1110010	34.8	100.0	65.2	73.6	0	0	0	2	9	23	11	9	19	8	19	19.93
125	s1101011	39.4	100.0	60.6	73.6	0	0	0	2	7	26	10	9	20	8	18	19.44
126	s1101010	34.8	100.0	65.2	73.6	0	0	0	3	8	23	10	10	19	8	19	20.02
127	s1111011	38.9	100.0	61.1	73.6	0	0	0	2	7	25	12	9	19	8	18	19.38
128	s1111010	34.8	100.0	65.2	73.5	0	0	0	3	8	23	11	9	19	8	19	19.97
	AVE	25.1	100.0	74.9	76.5	0.0	0.5	1.5	3.9	7.5	13.8	11.4	10.1	14.2	9.9	27.4	21.57

Ranked configuration analysis

CONFIGURATIONS
/hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile.ranked

NUM	CONFIG sTLSPCFU	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION 1	2	3	4	5	6	7	8	9	10	++	STDEV
1	s1010101	2	124	122	46.5	41	17	15	9	6	4	2	0	0	1	7	32.85
2	s1000101	2	124	122	46.6	41	17	14	9	5	6	2	0	0	1	7	32.72
3	s1010100	2	124	122	49.1	45	15	10	8	5	5	1	2	0	1	10	34.04
4	s1000100	2	124	122	49.1	45	15	10	8	4	6	1	2	0	1	10	34.02
5	s1011101	2	124	122	50.8	49	14	8	10	3	1	5	2	1	2	5	38.14
6	s1001101	2	124	122	51.1	48	15	7	10	3	2	5	2	1	2	5	37.93
7	s1011100	1	124	123	52.3	47	14	6	9	4	1	6	3	1	1	9	36.74
8	s1001100	1	124	123	52.3	46	15	6	9	4	1	6	3	1	1	9	36.69
9	s1010001	4	116	112	55.1	34	18	12	10	3	2	4	1	3	0	13	25.56
10	s1000001	4	116	112	55.2	34	18	12	10	3	2	4	1	2	1	13	25.58
11	s1010111	2	124	122	56.1	40	11	14	10	6	2	4	1	1	1	13	36.58
12	s1011111	2	124	122	56.3	40	11	14	10	6	2	3	1	2	1	13	36.69
13	s1000111	4	124	120	56.4	38	11	15	11	6	2	4	1	1	1	13	36.33
14	s1010000	4	127	123	56.4	38	15	11	9	4	2	2	2	1	0	19	28.58
15	s1000000	4	127	123	56.4	38	15	11	9	4	2	2	2	1	0	19	28.58
16	s1001111	4	124	120	56.6	38	11	15	11	6	2	3	1	2	1	13	36.44
17	s1110101	4	125	121	57.3	31	18	14	13	4	3	2	1	3	3	10	40.38
18	s1011001	8	124	116	57.5	33	12	16	8	5	5	2	2	3	0	15	26.24
19	s1001001	8	124	116	57.6	33	12	16	8	5	5	2	2	2	1	15	26.23
20	s1010110	3	125	122	57.7	42	11	9	8	6	4	4	2	1	1	13	35.01
21	s1011110	3	125	122	57.7	42	11	9	8	6	4	3	3	1	1	13	35.07
22	s1000110	4	127	123	57.8	41	11	10	8	6	4	4	2	1	1	13	34.95
23	s1001110	4	127	123	57.8	41	11	10	8	6	4	3	3	1	1	13	35.01
24	s1110100	4	125	121	58.0	32	16	11	12	4	4	2	3	5	1	12	38.30
25	s1100101	4	124	120	58.2	31	19	9	13	3	7	3	0	3	4	9	39.33
26	s1100100	4	124	120	58.6	32	17	8	12	3	6	3	2	6	2	10	37.44
27	s1011000	8	124	116	59.1	36	11	14	7	6	6	0	3	1	0	19	27.82
28	s1001000	8	124	116	59.1	36	11	14	7	6	6	0	3	1	0	19	27.82
29	s0001101	2	126	124	60.3	45	6	7	13	6	3	4	1	4	1	10	32.40
30	s0011101	2	126	124	60.3	45	6	7	13	6	3	4	1	4	1	10	32.40
31	s1100001	8	116	108	60.6	23	24	8	12	4	8	3	3	2	1	13	22.56
32	s1110001	8	120	112	60.8	23	24	11	8	5	6	3	5	2	1	14	24.10
33	s1111100	1	127	126	61.3	35	14	14	8	5	2	2	6	1	2	11	41.58
34	s1111101	1	127	126	61.7	39	13	15	8	4	1	1	4	2	1	12	44.14
35	s1101100	4	127	123	62.0	33	15	13	8	6	1	2	7	0	3	14	40.95
36	s1100000	8	124	116	62.2	25	22	10	9	4	8	2	4	1	0	18	23.91
37	s0000101	12	126	114	62.4	43	6	6	13	5	3	5	1	4	2	12	29.84
38	s0010101	12	126	114	62.4	43	6	6	13	5	3	5	1	4	2	12	29.84
39	s1110000	8	124	116	62.6	25	22	12	6	5	7	2	5	0	1	19	25.08
40	s1101101	2	127	125	62.7	35	15	13	9	4	2	2	4	0	3	15	43.27
41	s0000111	8	120	112	63.0	43	6	7	11	4	3	4	2	4	3	14	30.25
42	s0010111	8	120	112	63.0	43	6	7	11	4	3	4	2	4	3	14	30.25
43	s0011111	8	124	116	63.0	43	6	7	11	4	3	4	1	5	3	14	30.40
44	s0001111	8	124	116	63.0	43	6	7	11	4	3	4	1	5	3	14	30.40
45	s0001001	16	122	106	64.7	43	7	5	11	2	3	4	1	4	1	20	28.80
46	s0011001	16	122	106	64.7	43	7	5	11	2	3	4	1	4	1	20	28.80
47	s1101001	8	126	118	65.1	23	18	14	11	1	5	4	2	3	3	19	28.51
48	s1010010	6	121	115	65.2	35	6	14	6	11	2	1	5	2	2	17	29.53
49	s1010011	12	120	108	65.2	31	7	16	6	10	4	2	5	2	2	15	27.60
50	s1000010	6	123	117	65.2	35	6	14	6	11	2	1	4	3	2	17	29.50

continued below

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
51	s1000011	12	120	108	65.4	31	7	16	6	10	4	1	5	3	2	15	27.53
52	s1111001	4	126	122	65.4	24	17	14	10	2	4	4	4	4	3	17	30.02
53	s0010001	16	122	106	65.5	43	8	4	12	2	3	4	2	4	1	17	28.97
54	s0000001	16	122	106	65.5	43	8	4	12	2	3	4	2	4	1	17	28.97
55	s1011010	6	124	118	65.7	35	6	14	5	11	3	1	4	3	2	17	30.10
56	s1001010	6	124	118	65.8	35	6	14	5	11	3	1	3	4	2	17	30.07
57	s1011011	12	124	112	65.8	31	7	16	5	10	4	3	4	3	2	15	28.34
58	s0110101	1	124	123	65.9	33	15	8	11	6	2	3	6	1	4	12	33.50
59	s1001011	12	124	112	65.9	31	7	16	5	10	4	2	4	4	2	15	28.27
60	s0000011	16	112	96	66.0	43	6	5	11	2	5	4	2	4	1	17	28.14
61	s0010011	16	112	96	66.0	43	6	5	11	2	5	4	2	4	1	17	28.14
62	s0011011	16	112	96	66.1	43	6	5	11	2	4	5	2	4	1	17	28.25
63	s0001011	16	112	96	66.1	43	6	5	11	2	4	5	2	4	1	17	28.25
64	s0111101	2	127	125	66.2	36	11	11	10	8	3	3	3	0	3	14	35.84
65	s1110111	2	126	124	66.2	31	10	16	8	9	4	2	2	3	1	17	41.64
66	s1110110	4	126	122	66.3	30	10	14	7	10	7	3	0	3	1	17	37.78
67	s1111111	2	126	124	66.3	31	10	16	8	9	4	1	2	4	1	17	41.74
68	s1111110	4	126	122	66.3	30	10	14	7	10	7	2	1	3	1	17	37.78
69	s0100101	1	124	123	66.4	33	16	6	12	6	2	2	7	1	4	12	33.23
70	s0101101	2	127	125	66.5	35	12	9	11	9	3	3	3	0	3	14	35.68
71	s0010010	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
72	s0010100	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
73	s0010110	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
74	s0011010	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
75	s0011100	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
76	s0011000	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
77	s0011110	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
78	s0010000	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
79	s0000000	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
80	s0001010	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
81	s0000010	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
82	s0001000	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
83	s0000110	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
84	s0000100	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
85	s0001110	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
86	s0001100	16	120	104	66.7	46	5	4	12	3	3	3	2	4	1	17	32.04
87	s1101000	8	126	118	66.7	25	15	13	10	3	6	3	2	3	3	21	26.47
88	s1100110	4	127	123	67.0	29	9	14	5	12	6	6	0	2	3	17	36.02
89	s1111000	8	126	118	67.0	25	15	14	8	4	5	3	4	3	3	20	27.90
90	s1101110	4	124	120	67.1	29	9	14	5	12	6	5	1	2	3	17	36.05
91	s1100111	4	127	123	67.2	29	9	15	9	10	3	6	1	2	2	18	39.62
92	s1101111	4	124	120	67.3	29	9	15	9	10	3	5	1	3	2	18	39.69
93	s0100001	4	122	118	67.5	28	20	4	12	7	7	3	0	2	0	19	20.77
94	s0110100	4	124	120	67.6	33	14	7	11	6	2	2	8	2	2	14	32.05
95	s0110001	4	122	118	67.8	28	20	4	12	9	4	3	1	2	0	19	21.58
96	s0110111	4	124	120	67.8	32	11	9	11	8	4	4	1	0	4	20	32.12
97	s0111111	4	124	120	67.9	32	11	9	11	8	4	4	0	1	4	20	32.25
98	s0100100	4	124	120	68.0	33	15	5	12	6	2	1	9	2	2	14	31.86
99	s0111100	2	124	122	68.3	36	10	9	8	6	4	5	4	0	3	18	33.24
100	s0101001	20	123	103	68.4	27	14	7	15	4	7	7	1	2	0	17	22.91

continued below

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEV2/scorefile.ranked																		
NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION												STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++		
101	s0100111	4	124	120	68.5	31	12	9	11	7	5	4	1	0	4	18		31.62
102	s0101111	4	124	120	68.6	31	12	9	11	7	5	4	0	1	4	18		31.75
103	s0101100	2	124	122	68.6	35	11	8	9	6	3	5	4	1	3	18		33.03
104	s0100000	4	122	118	68.6	30	18	5	10	7	7	3	1	1	1	18		22.13
105	s0111001	20	123	103	68.7	27	14	7	15	6	4	7	2	2	0	17		23.79
106	s0110000	4	122	118	69.0	30	18	5	10	8	5	3	2	2	0	18		22.74
107	s0110110	4	124	120	69.7	33	10	8	10	6	5	4	2	1	3	23		29.09
108	s0111110	4	124	120	69.7	33	10	8	10	6	5	4	2	1	3	23		29.09
109	s0101000	26	122	96	69.8	29	12	7	14	6	6	6	1	1	1	19		22.43
110	s0101110	4	124	120	70.2	32	11	8	10	6	5	4	2	0	4	21		28.76
111	s0100110	4	124	120	70.2	32	11	8	10	6	5	4	2	0	4	21		28.76
112	s0111000	26	122	96	70.2	29	12	7	14	7	4	6	2	2	0	19		23.08
113	s0100011	32	120	88	71.8	27	13	7	13	7	9	1	2	3	0	20		20.83
114	s0101011	32	120	88	71.9	27	13	7	13	7	8	2	2	3	0	20		20.95
115	s0110011	32	120	88	72.1	27	13	7	13	9	6	1	3	3	0	20		21.58
116	s0111011	32	120	88	72.2	27	13	7	13	9	5	2	3	3	0	20		21.70
117	s0101010	32	120	88	72.4	29	12	7	12	8	8	1	1	1	1	23		22.14
118	s0100010	32	120	88	72.4	29	12	7	12	8	8	1	1	1	1	23		22.14
119	s0111010	32	120	88	72.8	29	12	7	12	9	6	1	2	2	0	23		22.69
120	s0110010	32	120	88	72.8	29	12	7	12	9	6	1	2	2	0	23		22.69
121	s1110010	16	124	108	73.3	23	10	14	7	13	4	3	6	2	1	20		26.83
122	s1100010	16	124	108	73.4	23	10	13	7	14	4	4	5	2	0	21		25.17
123	s1110011	16	124	108	73.6	21	11	15	7	11	4	3	8	3	3	16		27.21
124	s1100011	16	124	108	73.6	21	11	13	9	11	4	3	8	4	1	17		25.41
125	s1111010	16	124	108	73.7	23	10	14	6	13	5	3	5	3	1	20		27.32
126	s1101010	16	124	108	73.8	23	10	13	6	14	5	4	4	3	0	21		25.67
127	s1111011	16	124	108	74.0	21	11	15	6	11	4	4	7	4	3	16		27.82
128	s1101011	16	124	108	74.1	21	11	13	8	11	4	4	7	5	1	17		26.02
	AVE	10.3	123.3	113.0	64.5	35	11	9	9	6	4	3	2	2	1	16		30.66

Components

Numeric (left) and ranked (right) component analysis

`sT------`	1.262345	`sT------`	5.606563
`s-L-----`	-1.426397	`s-L-----`	-6.387813
`s--S----`	0.026936	`s--S----`	0.134219
`s---P---`	-0.222732	`s---P---`	-1.016094
`s----C--`	1.765016	`s----C--`	4.349687
`s-----F-`	-1.809405	`s-----F-`	-4.894063
`s------U`	0.807474	`s------U`	1.428437

C.4 Development corpus results III

This section presents the results tables using both the modified LAM grammar and the CRaP filter.

Documents

Numeric document analysis

DOCUMENTS /hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[|]	[?]
00	57.9	96.7	82.5	80.6	64.8	64.7	79.3	78.7	84.2	83.4	62.1	62.1		s01-0101/2	60.0	44.5
01	53.7	79.2	58.4	58.4	58.4	58.4	58.4	58.4	79.2	79.2	58.4	58.4			53.7	48.1
02	38.9	92.1	66.8	66.8	66.8	66.8	47.9	47.9	92.1	92.1	43.7	43.7			41.1	57.9
03	66.8	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3			78.1	28.9
04	48.4	90.2	77.8	77.8	56.2	56.2	77.8	77.8	77.8	77.8	56.2	56.2		s1---11-/16	69.3	39.9
05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.5
06	46.2	99.1	91.7	91.7	99.1	99.1	91.7	91.7	57.8	57.8	99.1	99.1			72.9	30.8
07	47.1	100.0	88.2	88.2	88.2	88.2	88.2	88.2	88.2	88.2	88.2	88.2		s00-----/32	100.0	18.4
08	85.0	100.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0		s1---0--/32	96.0	16.2
09	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	28.8
10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.9
11	25.7	100.0	59.2	59.2	54.9	54.9	59.2	59.2	25.7	37.8	54.9	54.9		s00-----/32	100.0	10.5
12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	39.3
13	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.6	82.6			82.6	25.8
14	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0	87.0			87.0	23.3
15	89.5	100.0	97.4	97.4	94.9	94.9	97.4	97.4	92.5	92.5	94.9	94.9		s01--1--/16	89.5	20.9
16	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.5
17	42.6	100.0	86.3	86.3	96.4	96.4	86.3	86.3	42.6	65.3	96.4	96.4		s10-----/24	96.4	16.7
18	50.0	64.0	64.0	64.0	60.3	60.3	64.0	64.0	58.1	64.0	60.3	60.3			51.5	47.8
19	85.4	100.0	85.4	85.4	96.9	96.9	85.4	85.4	85.4	85.4	96.9	96.9		s0------/64	100.0	18.9
20	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	21.3
21	70.8	100.0	96.2	96.2	96.2	96.2	96.2	96.2	70.8	70.8	96.2	96.2		s00-----/32	100.0	7.9
22	41.2	90.3	41.2	41.2	41.2	41.2	41.2	41.2	41.2	41.2	41.2	41.2		s0------/64	90.3	21.0
23	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	16.4
24	36.9	68.0	68.0	68.0	62.6	62.6	68.0	68.0	68.0	68.0	62.6	62.6			36.9	59.1
25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	38.2
26	47.5	67.1	67.1	66.9	63.8	63.6	67.1	66.9	55.2	53.8	63.8	63.6			54.5	48.4
27	47.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	63.7	63.7	100.0	100.0			100.0	13.2
28	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	22.7
29	37.5	88.0	47.4	47.4	78.6	78.6	47.4	47.4	37.5	40.9	78.6	78.6		s01-0-0-/8	75.2	29.0
30	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	12.2
31	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.0
32	34.8	69.3	69.3	69.3	59.5	59.5	69.3	69.3	69.1	68.6	59.5	59.5			34.8	61.1
33	37.6	92.4	84.3	84.3	70.0	70.0	84.3	84.3	76.2	76.2	70.0	70.0		s01-110-/4	37.6	55.7
34	76.1	83.0	79.7	79.7	79.7	79.7	79.7	79.7	79.7	79.7	79.7	79.7		s1---11-/16	78.2	31.8
35	45.0	62.6	62.6	62.6	45.0	45.0	62.6	62.6	62.6	62.6	45.0	45.0			54.4	50.9
36	56.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	67.3	67.3	100.0	100.0			100.0	40.0
37	53.3	100.0	53.3	53.3	100.0	100.0	53.3	53.3	53.3	53.3	100.0	100.0			74.3	33.3
38	28.6	73.7	73.7	73.7	61.6	61.6	73.7	73.7	73.7	73.7	61.6	61.6			28.6	70.2
39	47.3	66.4	66.4	66.4	60.2	60.2	66.4	66.4	66.4	66.4	60.2	60.2			47.3	55.0
40	20.0	98.2	87.1	87.1	87.1	87.1	87.1	87.1	83.4	79.1	87.1	87.1		s10--11-/8	20.0	74.8
41	54.1	100.0	100.0	100.0	81.4	81.4	54.1	54.1	97.7	97.7	54.1	54.1			54.1	47.9
42	36.4	94.1	72.1	72.1	94.1	94.1	72.1	72.1	36.4	52.9	94.1	94.1			90.9	17.2
43	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6	95.6			95.6	16.4
44	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.1
45	72.9	93.7	88.0	88.0	88.0	88.0	88.0	88.0	88.0	88.0	88.0	88.0		s1---11-/16	72.9	34.8
46	48.8	55.6	55.6	55.6	53.9	53.9	55.6	55.6	55.6	55.6	53.9	53.9			48.8	51.0
47	48.2	84.1	77.2	77.2	65.6	65.6	77.2	77.2	77.2	77.2	65.6	65.6		s-------/72	84.1	31.9
48	35.9	100.0	82.1	82.1	82.1	82.1	35.9	35.9	100.0	100.0	35.9	35.9			35.9	60.3
49	89.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	89.5	89.5	100.0	100.0			100.0	15.8
continued below																

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
50	89.8	100.0	89.8	89.8	89.8	89.8	89.8	89.8	100.0	100.0	89.8	89.8			89.8	30.0
51	84.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	84.6	84.6	100.0	100.0			100.0	14.1
52	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.7
53	50.8	88.1	88.1	88.1	84.9	84.9	61.8	61.8	88.1	88.1	61.8	61.8			61.8	39.2
54	97.3	97.3	97.3	97.3	97.3	97.3	97.3	97.3	97.3	97.3	97.3	97.3			97.3	11.8
55	67.8	100.0	82.0	82.0	82.0	82.0	100.0	100.0	82.0	82.0	100.0	100.0			100.0	9.6
56	32.3	78.7	44.7	44.7	44.7	44.7	44.7	44.7	78.7	76.1	44.7	44.7			32.3	62.8
57	52.9	75.7	64.3	64.3	64.3	64.3	64.3	64.3	64.3	64.3	64.3	64.3		s1---00-/16	60.0	43.8
58	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.2
59	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	10.8
60	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.6
61	69.0	100.0	99.0	99.0	100.0	100.0	99.0	99.0	99.0	99.0	100.0	100.0			69.0	35.0
62	37.5	64.1	59.5	54.8	59.5	54.8	42.5	42.5	59.5	54.8	42.5	42.5		s1-1-111/4	37.5	60.3
63	51.8	98.4	98.4	98.4	92.9	92.9	98.4	98.4	89.7	89.7	92.9	92.9			62.8	38.7
64	80.5	80.5	80.5	80.5	80.5	80.5	80.5	80.5	80.5	80.5	80.5	80.5			80.5	24.6
65	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2			75.2	35.2
66	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.6
67	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.5
68	53.3	100.0	81.9	81.9	100.0	100.0	81.9	81.9	74.3	74.3	100.0	100.0			90.5	22.9
69	41.6	73.2	73.2	73.2	66.4	66.4	73.2	73.2	51.3	48.4	66.4	66.4			64.1	39.3
70	53.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.1	95.1	100.0	100.0			100.0	8.6
71	49.4	94.7	78.0	78.0	74.5	74.5	78.0	78.0	66.3	62.9	74.5	74.5		s10--11-/8	58.5	42.8
72	46.2	84.6	84.6	84.6	84.6	84.6	84.6	84.6	71.8	71.8	84.6	84.6			53.8	42.3
73	25.3	70.6	54.9	54.9	54.9	54.9	54.9	54.9	64.1	56.8	54.9	54.9		s01-110-/4	25.3	67.6
74	44.9	100.0	64.7	64.7	73.5	73.5	89.0	89.0	51.5	51.5	100.0	100.0			77.2	36.8
75	54.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	72.7	72.7	100.0	100.0			83.0	29.4
76	46.2	84.6	70.2	70.2	64.0	64.0	70.2	70.2	70.2	70.2	64.0	64.0		s01--11-/8	56.0	47.1
77	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.6
78	24.6	92.9	92.9	92.9	83.1	83.1	70.8	70.8	92.9	88.3	71.1	71.1			24.6	66.5
79	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.8
80	41.0	100.0	76.9	76.9	76.9	76.9	76.9	76.9	64.1	64.1	76.9	76.9		s11--11-/8	48.7	41.0
81	84.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	84.6	84.6	100.0	100.0			100.0	25.6
82	31.8	100.0	100.0	100.0	87.9	87.9	72.7	72.7	100.0	92.4	56.1	56.1			31.8	66.7
83	39.6	85.8	46.7	46.7	46.7	46.7	46.7	46.7	46.7	46.7	46.7	46.7		s0------/64	85.8	35.3
84	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	18.7
85	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	22.3
86	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.2
87	57.4	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.9			57.4	44.2
88	16.7	86.8	20.0	20.0	23.4	23.4	20.0	20.0	86.8	84.3	23.4	23.4			16.7	74.9
89	24.2	98.8	98.8	98.8	42.5	42.5	71.0	71.0	98.8	73.4	28.1	28.1			24.2	68.3
90	47.3	100.0	63.7	63.7	63.7	63.7	63.7	63.7	63.7	63.7	63.7	63.7		s1---11-/16	47.3	51.6
91	46.1	74.6	65.0	65.0	70.1	70.1	65.0	65.0	56.6	58.4	70.1	70.1		s00-----/32	74.6	30.3
92	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8	62.8			62.8	40.0
93	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.2
94	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.3
95	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	26.4
96	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	15.5
97	35.5	70.9	67.4	67.4	62.4	62.4	70.9	70.9	62.8	62.5	65.7	65.7			36.0	62.6
98	38.2	100.0	56.4	56.4	56.4	56.4	56.4	56.4	100.0	100.0	56.4	56.4			38.2	56.4
99	31.2	92.8	92.8	92.8	82.2	82.2	78.6	78.6	92.8	92.8	65.2	65.2			31.2	60.9
AVE	65.5	92.0	84.1	84.0	83.0	83.0	82.0	82.0	81.7	81.7	81.4	81.4			75.5	33.2

KEY:
1: s1011101 2: s1001101 3: s1010101 4: s1000101 5: s1011100 6: s1001100 7: s1111101
8: s1101101 9: s1010100 10: s1000100 C:

Ranked document analysis

DOCUMENTS
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile.ranked

			OVERALL RANKED CONFIGURATIONS												
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[|]	[?]
00	1.5	127.5	77.0	78.0	29.0	36.0	39.0	40.0	93.0	93.0	97.5	97.5		s01-0101/2		
01	4.5	112.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	60.5	76.5	76.5		s11--10-/8		
02	5.5	127.5	42.5	42.5	42.5	42.5	77.5	77.5	101.5	101.5	5.5	5.5				
03	8.5	120.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	120.5	120.5				
04	8.5	124.5	96.5	96.5	20.5	20.5	20.5	20.5	96.5	96.5	8.5	8.5				
05	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
06	2.5	124.5	2.5	2.5	6.5	6.5	6.5	6.5	2.5	2.5	76.5	76.5				
07	16.5	124.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	124.5	124.5		s00-----/32		
08	16.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	104.5	104.5		s1---0--/32		
09	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
10	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
11	16.5	127.5	73.5	73.5	66.5	66.5	66.5	66.5	73.5	73.5	120.5	114.5		s00-----/32		
12	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
13	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
14	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
15	8.5	112.5	26.5	26.5	18.5	18.5	18.5	18.5	26.5	26.5	36.5	36.5		s01--1--/16		
16	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
17	12.5	127.5	42.5	42.5	92.5	92.5	92.5	92.5	42.5	42.5	12.5	12.5				
18	6.5	120.5	14.5	14.5	6.5	6.5	6.5	6.5	14.5	14.5	34.5	34.5				
19	32.5	116.5	72.5	72.5	116.5	116.5	116.5	116.5	72.5	72.5	116.5	116.5		s0------/64		
20	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
21	16.5	126.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5	56.5		s00-----/32		
22	32.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	72.5	72.5		s0------/64		
23	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
24	4.5	112.5	12.5	12.5	4.5	4.5	4.5	4.5	12.5	12.5	24.5	24.5				
25	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
26	3.5	124.5	17.5	19.5	3.5	7.5	3.5	7.5	17.5	19.5	3.5	3.5				
27	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
28	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
29	4.5	127.5	34.5	34.5	122.5	122.5	122.5	122.5	34.5	34.5	12.5	12.5		s01-0-0-/8		
30	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
31	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
32	2.5	112.5	22.5	22.5	2.5	2.5	2.5	2.5	22.5	22.5	12.5	12.5				
33	2.5	112.5	14.5	14.5	6.5	6.5	6.5	6.5	14.5	14.5	46.5	46.5		s01-110-/4		
34	8.5	120.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	40.5	8.5	8.5				
35	12.5	124.5	124.5	124.5	12.5	12.5	12.5	12.5	124.5	124.5	12.5	12.5				
36	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
37	4.5	124.5	4.5	4.5	124.5	124.5	124.5	124.5	4.5	4.5	16.5	16.5				
38	4.5	96.5	12.5	12.5	4.5	4.5	4.5	4.5	12.5	12.5	24.5	24.5				
39	4.5	96.5	20.5	20.5	4.5	4.5	4.5	4.5	20.5	20.5	56.5	56.5				
40	4.5	112.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	4.5	4.5				
41	2.5	112.5	26.5	26.5	2.5	2.5	112.5	112.5	112.5	112.5	12.5	12.5				
42	2.5	127.5	2.5	2.5	110.5	110.5	110.5	110.5	2.5	2.5	8.5	8.5				
43	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
44	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
45	8.5	96.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	8.5	8.5				
46	4.5	96.5	12.5	12.5	4.5	4.5	4.5	4.5	12.5	12.5	24.5	24.5				
47	36.5	120.5	92.5	92.5	80.5	80.5	80.5	80.5	92.5	92.5	120.5	120.5		s-------/72		
48	8.5	112.5	28.5	28.5	28.5	28.5	112.5	112.5	112.5	112.5	28.5	28.5		s-1--1-1/16		
49	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				

continued below

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile.ranked

			OVERALL RANKED CONFIGURATIONS												
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[|]	[?]
50	16.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5		s-1--1--/32		
51	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
52	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
53	4.5	124.5	12.5	12.5	4.5	4.5	64.5	64.5	64.5	64.5	116.5	124.5				
54	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
55	32.5	120.5	104.5	104.5	104.5	104.5	32.5	32.5	32.5	32.5	120.5	120.5				
56	8.5	112.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5	72.5		s-11-1--/16		
57	8.5	120.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	104.5	104.5		s1---00-/16		
58	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
59	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
60	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
61	4.5	96.5	4.5	4.5	12.5	12.5	12.5	12.5	4.5	4.5	24.5	24.5				
62	2.5	112.5	14.5	30.5	14.5	30.5	72.5	72.5	72.5	72.5	2.5	2.5				
63	2.5	124.5	6.5	6.5	2.5	2.5	2.5	2.5	6.5	6.5	114.5	114.5				
64	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
65	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
66	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
67	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
68	4.5	116.5	4.5	4.5	60.5	60.5	60.5	60.5	4.5	4.5	76.5	76.5				
69	2.5	126.5	6.5	6.5	2.5	2.5	2.5	2.5	6.5	6.5	110.5	110.5				
70	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
71	4.5	124.5	22.5	22.5	18.5	18.5	18.5	18.5	22.5	22.5	4.5	4.5				
72	4.5	120.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	120.5	120.5				
73	2.5	112.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5	52.5	52.5		s01-110-/4		
74	6.5	127.5	49.5	49.5	78.0	78.0	14.5	14.5	6.5	6.5	49.5	49.5				
75	4.5	126.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	12.5	12.5				
76	4.5	124.5	36.5	36.5	28.5	28.5	28.5	28.5	36.5	36.5	16.5	16.5		s01--11-/8		
77	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
78	2.0	120.5	6.0	6.0	2.0	2.0	15.5	15.5	13.5	13.5	24.5	24.5				
79	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
80	4.5	122.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5	20.5		s11--11-/8		
81	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
82	4.0	120.5	20.5	20.5	4.0	4.0	68.5	68.5	90.5	90.5	20.5	20.5				
83	32.5	120.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	72.5	72.5		s0------/64		
84	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
85	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
86	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
87	8.5	96.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	44.5	44.5				
88	1.5	112.5	66.5	66.5	82.5	82.5	82.5	82.5	66.5	66.5	82.5	82.5		s111110-/2		
89	2.0	120.5	83.5	83.5	2.0	2.0	14.5	14.5	109.5	109.5	36.0	49.0				
90	8.5	88.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	8.5	8.5				
91	16.5	127.5	38.5	38.5	62.5	62.5	62.5	62.5	38.5	38.5	98.5	94.5		s00-----/32		
92	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
93	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
94	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
95	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
96	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
97	1.5	126.5	19.5	19.5	3.5	3.5	1.5	1.5	9.5	9.5	6.5	6.5				
98	4.5	112.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5	52.5		s-1-110-/8		
99	2.5	112.5	6.5	6.5	2.5	2.5	10.5	10.5	54.5	54.5	18.5	18.5				
AVE	-	-	47.3	47.5	47.5	47.8	50.7	50.8	51.1	51.2	54.6	54.7			-	-

KEY:
1: s1010101 2: s1000101 3: s1011101 4: s1001101 5: s1011100 6: s1001100 7: s1010100
8: s1000100 9: s1010111 10: s1011111 C:

Configurations

Numeric configuration analysis

CONFIGURATIONS																	
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
1	s1011101	20.0	100.0	80.0	84.1	0	1	0	0	4	7	13	10	17	12	36	17.99
2	s1001101	20.0	100.0	80.0	84.0	0	1	0	0	4	7	13	10	17	12	36	18.07
3	s1010101	23.4	100.0	76.6	83.0	0	0	1	0	5	8	14	8	16	11	37	18.97
4	s1000101	23.4	100.0	76.6	83.0	0	0	1	0	5	8	14	8	16	11	37	19.04
5	s1011100	20.0	100.0	80.0	82.0	0	1	0	1	6	7	11	16	14	9	35	19.31
6	s1001100	20.0	100.0	80.0	82.0	0	1	0	1	6	7	11	16	14	9	35	19.31
7	s1111101	25.7	100.0	74.3	81.7	0	0	1	2	3	9	13	13	16	12	31	18.92
8	s1101101	37.8	100.0	62.2	81.7	0	0	0	1	4	9	14	15	16	11	30	17.85
9	s1010100	23.4	100.0	76.6	81.4	0	0	2	1	6	9	15	8	9	11	39	21.14
10	s1000100	23.4	100.0	76.6	81.4	0	0	2	1	6	9	15	8	9	11	39	21.14
11	s1100101	36.4	100.0	63.6	81.2	0	0	0	1	4	8	18	12	17	9	31	17.86
12	s1110101	29.9	100.0	70.1	81.1	0	0	1	0	4	9	18	10	18	9	31	18.15
13	s1010111	20.0	100.0	80.0	81.0	0	1	1	0	6	13	11	6	13	14	35	20.38
14	s1011111	20.0	100.0	80.0	81.0	0	1	0	1	6	13	11	6	13	14	35	20.38
15	s1000111	20.0	100.0	80.0	81.0	0	1	1	0	6	14	10	6	13	14	35	20.45
16	s1001111	20.0	100.0	80.0	81.0	0	1	0	1	6	14	10	6	13	14	35	20.45
17	s1101100	37.8	100.0	62.2	80.9	0	0	0	1	5	7	17	18	14	8	30	17.89
18	s1111100	25.7	100.0	74.3	80.6	0	0	1	2	4	7	16	17	15	7	31	18.92
19	s1100100	36.4	100.0	63.6	80.3	0	0	0	1	5	6	23	12	13	9	31	18.31
20	s0111101	28.6	100.0	71.4	80.2	0	0	1	0	9	11	10	13	15	8	33	19.48
21	s1110100	29.9	100.0	70.1	80.2	0	0	1	0	5	7	24	9	14	9	31	18.67
22	s1101111	44.9	100.0	55.1	80.0	0	0	0	0	7	17	10	8	18	9	31	18.80
23	s1100111	42.2	100.0	57.8	80.0	0	0	0	0	7	17	10	8	18	9	31	18.83
24	s1111111	28.5	100.0	71.5	79.8	0	0	1	0	7	11	17	6	17	10	31	19.29
25	s0101101	28.6	100.0	71.4	79.8	0	0	1	1	9	9	11	14	14	9	32	19.81
26	s1110111	26.9	100.0	73.1	79.8	0	0	1	0	7	11	17	6	17	10	31	19.33
27	s1010110	20.0	100.0	80.0	79.7	0	1	1	3	7	14	8	6	11	12	37	22.05
28	s1000110	20.0	100.0	80.0	79.7	0	1	1	3	7	14	8	6	11	12	37	22.05
29	s1011110	20.0	100.0	80.0	79.7	0	1	0	4	7	14	8	6	11	12	37	22.09
30	s1001110	20.0	100.0	80.0	79.7	0	1	0	4	7	14	8	6	11	12	37	22.09
31	s0001101	16.7	100.0	83.3	79.4	0	1	3	5	4	11	6	11	13	8	38	23.12
32	s0011101	16.7	100.0	83.3	79.4	0	1	3	5	4	11	6	11	13	8	38	23.12
33	s1011001	20.0	100.0	80.0	79.3	0	1	0	2	8	11	13	9	14	7	35	20.90
34	s1001001	20.0	100.0	80.0	79.3	0	1	0	2	8	11	13	9	14	7	35	20.91
35	s1101110	38.5	100.0	61.5	79.2	0	0	0	1	9	13	10	12	16	8	31	19.40
36	s1100110	38.5	100.0	61.5	79.2	0	0	0	1	9	13	10	12	16	8	31	19.43
37	s0110101	28.6	100.0	71.4	79.0	0	0	1	0	8	13	15	9	16	7	31	19.59
38	s1111110	28.5	100.0	71.5	78.9	0	0	1	1	8	11	14	9	16	9	31	19.94
39	s1110110	26.9	100.0	73.1	78.9	0	0	1	1	8	11	14	9	16	9	31	19.97
40	s1100001	37.0	100.0	63.0	78.8	0	0	0	1	9	12	13	9	18	10	28	19.34
41	s1110001	37.5	100.0	62.5	78.8	0	0	0	1	9	12	13	10	18	9	28	19.25
42	s0100101	28.6	100.0	71.4	78.7	0	0	1	1	8	11	16	9	16	7	31	20.01
43	s1111001	40.7	100.0	59.3	78.7	0	0	0	0	7	17	13	9	16	9	29	19.14
44	s1101001	40.7	100.0	59.3	78.7	0	0	0	0	7	19	10	10	15	11	28	19.14
45	s1010001	20.0	100.0	80.0	78.7	0	1	0	3	10	11	11	6	15	7	36	22.04
46	s1000001	20.0	100.0	80.0	78.6	0	1	0	3	10	11	11	6	15	7	36	22.05
47	s0111100	28.6	100.0	71.4	78.5	0	0	1	4	8	9	12	14	12	7	33	20.98
48	s1001000	20.0	100.0	80.0	78.4	0	1	0	6	6	12	10	8	15	6	36	22.32
49	s1011000	20.0	100.0	80.0	78.4	0	1	0	6	6	12	10	8	15	6	36	22.32
50	s0110111	28.6	100.0	71.4	78.4	0	0	1	0	7	16	13	13	13	6	31	19.44
continued below																	

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile

NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
51	s0111111	28.6	100.0	71.4	78.4	0	0	1	0	7	16	13	13	13	6	31	19.46
52	s1010000	20.0	100.0	80.0	78.3	0	1	1	5	8	13	8	6	14	6	38	23.09
53	s1000000	20.0	100.0	80.0	78.3	0	1	1	5	8	13	8	6	14	6	38	23.09
54	s0101100	28.6	100.0	71.4	78.2	0	0	1	5	8	7	13	14	12	8	32	21.27
55	s1100000	34.8	100.0	65.2	78.2	0	0	0	3	8	12	13	8	18	9	29	20.15
56	s1110000	34.8	100.0	65.2	78.1	0	0	0	3	8	12	12	10	18	8	29	20.07
57	s0100111	28.6	100.0	71.4	78.1	0	0	1	0	8	15	13	13	13	6	31	19.68
58	s0101111	28.6	100.0	71.4	78.1	0	0	1	0	8	15	13	13	13	6	31	19.70
59	s0010101	16.7	100.0	83.3	78.0	0	1	3	5	5	11	9	10	12	7	37	23.46
60	s0000101	16.7	100.0	83.3	78.0	0	1	3	5	5	11	9	10	12	7	37	23.46
61	s0000111	16.7	100.0	83.3	77.9	0	1	3	4	5	12	11	10	9	8	37	23.34
62	s0010111	16.7	100.0	83.3	77.9	0	1	3	4	5	12	11	10	9	8	37	23.34
63	s0011111	16.7	100.0	83.3	77.8	0	1	3	4	5	12	11	10	9	8	37	23.35
64	s0001111	16.7	100.0	83.3	77.8	0	1	3	4	5	12	11	10	9	8	37	23.35
65	s1101000	34.8	100.0	65.2	77.8	0	0	0	1	8	17	11	9	16	9	29	20.05
66	s1111000	34.8	100.0	65.2	77.8	0	0	0	1	8	17	12	8	17	8	29	20.04
67	s0110100	28.6	100.0	71.4	77.7	0	0	1	4	8	9	17	10	13	7	31	20.97
68	s0100100	28.6	100.0	71.4	77.5	0	0	1	5	9	6	18	10	13	7	31	21.32
69	s0111001	27.1	100.0	72.9	77.2	0	0	2	3	8	12	12	11	15	9	28	21.09
70	s0101001	28.6	100.0	71.4	77.2	0	0	2	3	8	13	11	11	15	9	28	21.08
71	s0110001	27.1	100.0	72.9	77.0	0	0	2	5	8	11	12	8	16	10	28	21.84
72	s0100001	28.6	100.0	71.4	77.0	0	0	2	5	8	11	12	8	16	10	28	21.82
73	s0111110	28.6	100.0	71.4	76.9	0	0	1	5	7	11	15	12	12	6	31	21.08
74	s0110110	28.6	100.0	71.4	76.9	0	0	1	5	7	11	15	12	12	6	31	21.08
75	s0100110	28.6	100.0	71.4	76.7	0	0	1	5	8	10	15	12	12	6	31	21.28
76	s0101110	28.6	100.0	71.4	76.7	0	0	1	5	8	10	15	12	12	6	31	21.28
77	s0011001	16.7	100.0	83.3	76.7	0	1	3	6	9	10	9	9	8	8	37	24.47
78	s0001001	16.7	100.0	83.3	76.7	0	1	3	6	9	10	9	9	8	8	37	24.47
79	s0110000	27.1	100.0	72.9	76.4	0	0	3	6	7	10	13	7	16	9	29	22.67
80	s0100000	28.0	100.0	72.0	76.4	0	0	3	6	7	10	13	7	16	9	29	22.66
81	s1010011	20.0	100.0	80.0	76.3	0	1	0	3	7	23	9	3	14	5	35	22.56
82	s1000011	20.0	100.0	80.0	76.3	0	1	0	3	7	23	9	3	14	5	35	22.57
83	s0111000	27.1	100.0	72.9	76.2	0	0	3	5	8	9	14	10	14	8	29	22.28
84	s0101000	28.0	100.0	72.0	76.2	0	0	3	5	8	9	14	10	14	8	29	22.27
85	s1011011	20.0	100.0	80.0	76.2	0	1	0	3	8	22	9	3	14	5	35	22.63
86	s1001011	20.0	100.0	80.0	76.2	0	1	0	3	8	22	9	3	14	5	35	22.64
87	s1010010	20.0	100.0	80.0	76.2	0	1	0	5	6	23	8	3	12	5	37	23.35
88	s1000010	20.0	100.0	80.0	76.2	0	1	0	5	6	23	8	3	12	5	37	23.35
89	s1011010	20.0	100.0	80.0	76.1	0	1	0	5	7	22	8	3	12	5	37	23.40
90	s1001010	20.0	100.0	80.0	76.1	0	1	0	5	7	22	8	3	12	5	37	23.40
91	s0010001	16.7	100.0	83.3	76.0	0	1	3	8	9	8	10	8	8	8	37	25.20
92	s0000001	16.7	100.0	83.3	76.0	0	1	3	8	9	8	10	8	8	8	37	25.20
93	s0000011	16.7	100.0	83.3	76.0	0	1	3	6	11	9	9	8	9	7	37	25.02
94	s0010011	16.7	100.0	83.3	76.0	0	1	3	6	11	9	9	8	9	7	37	25.02
95	s0001011	16.7	100.0	83.3	76.0	0	1	3	6	11	9	9	8	9	7	37	25.04
96	s0011011	16.7	100.0	83.3	76.0	0	1	3	6	11	9	9	8	9	7	37	25.04
97	s1100011	39.4	100.0	60.6	75.8	0	0	0	2	7	25	8	7	17	6	28	20.54
98	s1101011	39.4	100.0	60.6	75.8	0	0	0	2	7	25	8	7	17	6	28	20.61
99	s1110011	39.6	100.0	60.4	75.8	0	0	0	1	8	24	10	7	17	5	28	20.45
100	s1111011	38.9	100.0	61.1	75.7	0	0	0	2	7	24	10	7	17	5	28	20.53

continued below

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile																		
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
101	s1100010	34.8	100.0	65.2	75.6	0	0	0	3	8	22	9	8	15	6	29	21.09
102	s1101010	34.8	100.0	65.2	75.5	0	0	0	3	8	22	9	8	15	6	29	21.13
103	s0011000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
104	s0001100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
105	s0011100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
106	s0000000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
107	s0010000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
108	s0010010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
109	s0011010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
110	s0000010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
111	s0010100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
112	s0010110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
113	s0000110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
114	s0001110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
115	s0000100	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
116	s0001010	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
117	s0011110	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
118	s0001000	16.7	100.0	83.3	75.5	0	1	5	10	5	9	8	9	8	7	38	26.23
119	s1110010	34.8	100.0	65.2	75.5	0	0	0	2	9	22	10	7	16	5	29	21.02
120	s1111010	34.8	100.0	65.2	75.4	0	0	0	3	8	22	10	7	16	5	29	21.07
121	s0110011	28.6	100.0	71.4	75.4	0	0	1	4	12	14	10	10	14	7	28	21.87
122	s0111011	28.6	100.0	71.4	75.4	0	0	1	4	12	14	10	10	14	7	28	21.89
123	s0100011	28.6	100.0	71.4	75.3	0	0	1	4	12	14	10	10	14	7	28	21.86
124	s0101011	28.6	100.0	71.4	75.3	0	0	1	4	12	14	10	10	14	7	28	21.89
125	s0111010	28.0	100.0	72.0	74.8	0	0	2	7	9	11	14	8	13	7	29	22.75
126	s0110010	28.0	100.0	72.0	74.8	0	0	2	7	9	11	14	8	13	7	29	22.75
127	s0100010	28.0	100.0	72.0	74.8	0	0	2	7	9	11	14	8	13	7	29	22.75
128	s0101010	28.0	100.0	72.0	74.8	0	0	2	7	9	11	14	8	13	7	29	22.75
	AVE	25.2	100.0	74.8	77.9	0.0	0.5	1.5	3.9	7.1	12.4	11.3	9.0	13.1	7.9	33.4	21.85

Ranked configuration analysis

CONFIGURATIONS																	
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile.ranked																	
NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
1	s1010101	2	124	122	47.3	49	17	10	8	5	3	2	0	0	1	7	29.70
2	s1000101	2	124	122	47.5	49	17	9	8	4	5	2	0	0	1	7	29.56
3	s1011101	2	124	122	47.5	60	14	6	7	2	1	2	1	1	2	4	33.37
4	s1001101	2	124	122	47.8	59	15	5	7	2	2	2	1	1	2	4	33.16
5	s1011100	1	124	123	50.7	56	14	5	6	3	1	3	2	1	1	9	32.88
6	s1001100	1	124	123	50.8	55	15	5	6	3	1	3	2	1	1	9	32.82
7	s1010100	2	124	122	51.1	51	15	6	8	4	4	1	2	0	1	10	31.32
8	s1000100	2	124	122	51.2	51	15	6	8	3	5	1	2	0	1	10	31.30
9	s1010111	2	124	122	54.6	51	9	10	10	5	1	2	1	1	1	11	33.38
10	s1011111	2	124	122	54.7	51	9	10	10	5	1	2	0	2	1	11	33.32
11	s1110101	4	125	121	54.7	42	17	11	10	3	3	2	1	2	2	9	35.76
12	s1111101	1	127	126	54.7	53	13	12	6	2	0	0	3	2	0	9	38.66
13	s1000111	4	124	120	54.8	49	9	11	11	5	1	2	1	1	1	11	33.11
14	s1001111	4	124	120	55.0	49	9	11	11	5	1	2	0	2	1	11	33.06
15	s1100101	4	124	120	55.6	42	18	6	11	2	6	3	0	2	3	8	34.52
16	s1111100	1	127	126	55.6	48	14	12	6	3	1	0	5	1	1	9	36.55
17	s1101101	2	126	124	55.7	49	15	10	7	2	1	1	3	0	2	12	37.81
18	s1110100	4	125	121	56.3	42	15	9	9	3	4	2	3	3	1	11	33.89
19	s1101100	4	126	122	56.3	46	15	11	6	4	0	0	6	0	2	12	35.91
20	s1100100	4	124	120	56.8	42	16	6	10	2	5	3	2	4	2	9	32.83
21	s1011001	8	124	116	56.9	44	7	14	7	4	4	2	2	3	0	13	23.91
22	s1001001	8	124	116	56.9	44	7	14	7	4	4	2	2	2	1	13	23.90
23	s1011110	3	125	122	57.8	51	9	6	8	5	3	2	2	1	1	13	32.08
24	s1010110	3	125	122	57.8	51	9	6	8	5	3	2	2	1	1	13	32.19
25	s1001110	4	127	123	57.9	50	9	7	8	5	3	2	2	1	1	13	32.01
26	s1000110	4	127	123	57.9	50	9	7	8	5	3	2	2	1	1	13	32.12
27	s1010001	4	116	112	58.0	44	8	12	10	3	2	4	1	3	0	13	23.54
28	s1000001	4	116	112	58.0	44	8	12	10	3	2	4	1	2	1	13	23.54
29	s1011000	8	124	116	59.8	45	7	12	6	5	5	0	3	1	0	19	25.63
30	s1001000	8	124	116	59.8	45	7	12	6	5	5	0	3	1	0	19	25.63
31	s1000000	4	127	123	60.0	46	7	11	9	4	2	2	2	1	0	19	26.34
32	s1010000	4	127	123	60.0	46	7	11	9	4	2	2	2	1	0	19	26.34
33	s1110111	2	126	124	60.5	45	10	10	8	8	2	0	2	3	1	13	37.48
34	s1111111	2	126	124	60.5	45	10	10	8	8	2	0	1	4	1	13	37.46
35	s1100001	8	116	108	60.6	36	15	8	12	2	7	3	2	2	1	13	19.72
36	s1110001	8	120	112	61.0	36	15	10	8	4	5	3	4	2	1	14	21.68
37	s1101001	8	126	118	61.1	37	14	12	9	0	4	3	2	2	2	18	25.63
38	s1100111	4	127	123	61.3	43	9	10	9	8	1	4	1	2	2	14	35.24
39	s1111001	4	126	122	61.4	38	13	12	8	1	3	3	4	2	3	16	27.39
40	s1101111	4	124	120	61.4	43	9	10	9	8	1	4	0	3	2	14	35.20
41	s1111110	4	126	122	61.8	43	10	9	7	9	5	0	0	3	1	14	33.84
42	s1110110	4	126	122	61.9	43	10	9	7	9	5	0	0	3	1	14	34.00
43	s1101110	4	124	120	62.5	42	9	10	5	10	4	3	0	2	3	14	31.87
44	s1100110	4	127	123	62.5	42	9	10	5	10	4	3	0	2	3	14	31.99
45	s1100000	8	124	116	62.9	37	14	9	9	3	7	2	3	1	0	18	21.03
46	s0011101	2	126	124	63.0	50	5	5	12	6	3	4	1	4	1	9	29.73
47	s0001101	2	126	124	63.0	50	5	5	12	6	3	4	1	4	1	9	29.73
48	s1110000	8	124	116	63.4	37	14	10	6	5	6	2	4	0	1	19	22.53
49	s0111101	2	127	125	63.6	47	10	8	9	7	2	3	2	0	2	11	32.57
50	s1101000	8	126	118	63.8	38	12	10	8	2	5	2	2	2	2	21	23.96
continued below																	

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
51	s0101101	2	127	125	63.9	46	11	6	10	8	2	3	2	0	2	11	32.41
52	s1111000	8	126	118	64.2	38	12	11	6	3	4	2	4	1	3	20	25.62
53	s0110101	1	124	123	64.6	44	12	6	8	6	2	3	5	1	3	10	31.20
54	s1010011	12	120	108	64.8	42	1	16	5	8	3	2	4	2	2	15	25.52
55	s1000011	12	120	108	64.9	42	1	16	5	8	3	1	4	3	2	15	25.45
56	s1011011	12	120	108	65.0	42	1	16	5	8	2	3	4	2	2	15	25.92
57	s0100101	1	124	123	65.2	44	13	4	9	6	2	2	6	1	3	10	30.92
58	s1001011	12	120	108	65.2	42	1	16	5	8	2	2	4	3	2	15	25.85
59	s1010010	6	121	115	65.7	44	1	14	5	9	2	1	4	2	2	17	27.01
60	s0110111	4	123	119	65.7	43	9	6	10	7	3	4	1	0	3	18	28.63
61	s0010101	12	126	114	65.7	48	5	4	11	5	3	5	1	4	2	12	27.06
62	s0000101	12	126	114	65.7	48	5	4	11	5	3	5	1	4	2	12	27.06
63	s1000010	6	123	117	65.7	44	1	14	5	9	2	1	3	3	2	17	26.97
64	s0111111	4	124	120	65.7	43	9	6	10	7	3	4	0	1	3	18	28.78
65	s0000111	8	120	112	65.8	48	5	5	10	4	3	4	2	4	3	12	27.22
66	s0010111	8	120	112	65.8	48	5	5	10	4	3	4	2	4	3	12	27.22
67	s0011111	8	124	116	65.9	48	5	5	10	4	3	4	1	5	3	12	27.38
68	s0001111	8	124	116	65.9	48	5	5	10	4	3	4	1	5	3	12	27.38
69	s1011010	6	121	115	65.9	44	1	14	5	9	2	1	4	2	2	17	27.24
70	s1001010	6	123	117	66.0	44	1	14	5	9	2	1	3	3	2	17	27.20
71	s0100111	4	123	119	66.3	42	10	6	10	6	4	4	1	0	3	16	28.12
72	s0101111	4	124	120	66.4	42	10	6	10	6	4	4	0	1	3	16	28.28
73	s0111100	2	124	122	67.0	46	9	7	7	5	3	4	3	0	2	16	30.47
74	s0110100	4	124	120	67.2	43	11	6	8	6	2	2	7	1	2	12	29.90
75	s0101100	2	124	122	67.3	45	10	6	8	5	2	4	3	1	2	16	30.25
76	s0100100	4	124	120	67.6	43	12	4	9	6	2	1	8	1	2	12	29.69
77	s0101001	20	123	103	67.7	38	11	5	12	4	6	7	0	2	0	16	20.46
78	s0001001	16	122	106	67.9	48	5	4	10	2	3	4	1	4	1	18	25.76
79	s0011001	16	122	106	67.9	48	5	4	10	2	3	4	1	4	1	18	25.76
80	s0111001	20	123	103	68.0	38	11	5	12	6	3	7	1	2	0	16	21.45
81	s0100001	4	122	118	68.5	39	14	3	10	5	7	3	0	2	0	19	19.50
82	s0110001	4	122	118	68.8	39	14	3	10	7	4	3	1	2	0	19	20.35
83	s0110110	4	120	116	68.9	43	8	6	9	5	4	3	2	1	2	22	26.10
84	s0111110	4	120	116	68.9	43	8	6	9	5	4	3	2	1	2	22	26.10
85	s0000001	16	122	106	69.3	48	5	3	11	2	3	4	2	4	1	17	25.85
86	s0010001	16	122	106	69.3	48	5	3	11	2	3	4	2	4	1	17	25.85
87	s0100110	4	120	116	69.3	42	9	6	9	5	4	3	2	0	3	20	25.74
88	s0101110	4	120	116	69.3	42	9	6	9	5	4	3	2	0	3	20	25.74
89	s0000011	16	112	96	69.5	48	4	4	10	2	4	4	2	4	1	17	25.08
90	s0010011	16	112	96	69.5	48	4	4	10	2	4	4	2	4	1	17	25.08
91	s0001011	16	112	96	69.6	48	4	4	10	2	3	5	2	4	1	17	25.19
92	s0011011	16	112	96	69.6	48	4	4	10	2	3	5	2	4	1	17	25.19
93	s1100011	16	124	108	70.0	35	7	12	7	8	3	3	5	4	1	17	24.25
94	s1110011	16	124	108	70.1	35	7	13	5	9	3	3	5	3	3	16	26.26
95	s1101011	16	124	108	70.2	35	7	12	7	8	2	4	5	4	1	17	24.59
96	s0100000	4	122	118	70.3	40	13	3	8	6	7	3	1	1	1	18	20.45
97	s1111011	16	124	108	70.3	35	7	13	5	9	2	4	5	3	3	16	26.60
98	s0101000	26	122	96	70.3	39	10	4	11	6	5	6	0	1	1	19	20.00
99	s0110000	4	122	118	70.7	40	13	3	8	7	5	3	2	2	0	18	21.08
100	s1100010	16	124	108	70.7	36	7	11	5	11	3	4	3	2	0	21	23.58

continued below

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/DEV3/scorefile.ranked																
NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION										STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
101	s0111000	26	122	96	70.7	39	10	4	11	7	3	6	1	2	0	19	20.71
102	s1110010	16	124	108	70.8	36	7	11	5	11	3	3	4	2	1	20	25.48
103	s1101010	16	124	108	70.8	36	7	11	5	11	3	4	3	2	0	21	23.79
104	s1111010	16	124	108	70.9	36	7	11	5	11	3	3	4	2	1	20	25.68
105	s0010110	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
106	s0011110	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
107	s0011100	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
108	s0011010	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
109	s0001010	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
110	s0000000	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
111	s0011000	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
112	s0010010	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
113	s0010000	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
114	s0001110	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
115	s0000100	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
116	s0001100	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
117	s0000010	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
118	s0001000	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
119	s0000110	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
120	s0010100	16	120	104	71.2	49	4	3	11	3	3	3	2	4	1	17	28.56
121	s0100011	32	120	88	72.2	38	9	5	10	6	9	1	1	3	0	20	19.18
122	s0101011	32	120	88	72.3	38	9	5	10	6	8	2	1	3	0	20	19.32
123	s0110011	32	120	88	72.5	38	9	5	10	8	6	1	2	3	0	20	19.99
124	s0111011	32	120	88	72.6	38	9	5	10	8	5	2	2	3	0	20	20.12
125	s0101010	32	120	88	73.7	39	9	4	9	7	8	1	1	1	1	23	20.02
126	s0100010	32	120	88	73.7	39	9	4	9	7	8	1	1	1	1	23	20.02
127	s0110010	32	120	88	74.0	39	9	4	9	8	6	1	2	2	0	23	20.60
128	s0111010	32	120	88	74.0	39	9	4	9	8	6	1	2	2	0	23	20.60
	AVE	10.3	123.1	112.8	64.5	44	8	7	8	5	3	2	2	2	1	15	27.72

Components

Numeric (left) and ranked (right) component analysis

sT------	2.347637	sT------	9.084062
s-L-----	-0.085065	s-L-----	-2.484062
s--S----	0.015399	s--S----	0.107969
s---P---	0.209383	s---P---	0.173906
s----C--	2.636099	s----C--	5.892188
s-----F-	-1.541117	s-----F-	-4.314063
s------U	1.117817	s------U	2.578438

C.5 Test corpus results

Documents

Numeric document analysis

DOCUMENTS /hame/jeremyc/phd/eval/sys/scores/TEST/scorefile																
			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
00	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	22.2
01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	17.5
02	42.3	71.8	43.6	43.6	43.6	43.6	43.6	43.6	43.6	43.6	47.4	44.9		s01--11-/8	53.8	43.6
03	46.9	71.5	57.0	57.0	57.0	57.0	71.5	71.5	71.5	71.5	63.2	63.2			60.4	40.4
04	93.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			93.8	13.7
05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	9.4
06	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0			91.0	14.4
07	42.5	78.6	63.5	63.5	63.5	63.5	71.9	71.9	71.9	71.9	78.6	78.6			44.9	53.8
08	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.5
09	47.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	9.9
10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.0
11	82.0	100.0	100.0	100.0	100.0	100.0	82.0	82.0	82.0	82.0	82.0	82.0			82.0	28.2
12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	33.3
13	59.5	100.0	89.5	89.5	89.5	89.5	89.5	89.5	89.5	89.5	89.5	89.5		s-1-111-/8	89.5	21.6
14	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.8
15	38.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	19.5
16	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8	94.8			94.8	16.2
17	47.4	70.5	57.7	57.7	57.7	57.7	57.7	57.7	57.7	57.7	70.5	70.5			53.8	44.9
18	40.9	100.0	91.6	91.6	91.6	91.6	91.6	91.6	91.6	91.6	76.8	91.6		s00-----/32	100.0	14.8
19	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	16.7
20	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	21.7
21	48.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.3
22	68.3	100.0	82.3	82.3	82.3	82.3	82.3	82.3	82.3	82.3	68.3	68.3		s0------/64	100.0	16.9
23	68.5	100.0	68.5	68.5	68.5	68.5	100.0	100.0	100.0	100.0	100.0	100.0			100.0	17.5
24	63.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.8
25	64.4	100.0	100.0	100.0	100.0	100.0	76.3	76.3	76.3	76.3	100.0	100.0			64.4	41.1
26	62.1	86.6	75.1	75.1	75.1	75.1	75.1	75.1	75.1	75.1	69.8	69.8		s10--0--/16	85.9	24.0
27	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	7.8
28	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.1
29	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.1
30	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	19.2
31	40.1	56.5	47.4	47.4	47.4	47.4	47.4	47.4	47.4	47.4	56.5	56.5			42.3	59.3
32	29.9	64.2	40.0	40.0	40.0	40.0	29.9	29.9	29.9	29.9	29.9	29.9		s---1101/8	29.9	63.4
33	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	10.2
34	81.6	87.5	84.9	84.9	84.9	84.9	84.9	84.9	84.9	84.9	81.6	81.6		s0------/64	87.5	26.6
35	53.3	100.0	57.1	57.1	57.1	57.1	100.0	100.0	100.0	100.0	100.0	100.0			100.0	18.1
36	19.3	84.7	81.1	80.0	81.1	80.0	58.1	58.1	58.1	58.1	79.2	73.9		s1011101/1	19.3	74.7
37	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.9
38	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.3
39	41.8	67.3	67.3	67.3	67.3	67.3	67.3	67.3	67.3	67.3	67.3	67.3			67.3	34.5
40	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.6
41	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.1
42	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	8.7
43	93.2	100.0	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6	93.2	93.2		s0------/64	100.0	16.2
44	48.3	85.2	52.9	52.9	52.9	52.9	52.9	52.9	52.9	52.9	48.3	50.8		s0------/64	85.2	26.2
45	75.2	86.7	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2	75.2		s-------/96	86.7	28.6
46	68.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	17.1
47	28.8	100.0	84.9	84.9	84.9	84.9	84.9	84.9	84.9	84.9	100.0	83.8			28.8	67.0
48	68.0	77.0	71.8	71.8	71.8	71.8	71.8	71.8	71.8	71.8	68.0	70.6		s1--100-/8	74.9	33.5
49	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	23.5
continued below																

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[]	[?]
50	46.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	46.7	47.5			60.0	45.0
51	77.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			77.9	39.0
52	94.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.6	100.0			94.6	18.6
53	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	16.7
54	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.0
55	47.6	82.7	82.7	82.7	82.7	82.7	82.7	82.7	82.7	82.7	58.4	58.4			54.5	44.6
56	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.9
57	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.3
58	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.8
59	39.4	54.5	45.9	45.9	45.9	45.9	39.4	39.4	39.4	39.4	39.4	39.4		s-0-1101/4	39.4	55.0
60	85.1	85.1	85.1	85.1	85.1	85.1	85.1	85.1	85.1	85.1	85.1	85.1			85.1	23.6
61	41.3	58.0	56.7	56.7	55.3	55.3	56.7	56.7	55.3	55.3	46.0	43.3		s11-111-/4	41.3	58.3
62	44.8	66.7	58.1	58.1	58.1	58.1	58.1	58.1	58.1	58.1	66.7	66.7			46.7	53.3
63	52.7	100.0	75.8	75.8	75.8	75.8	75.8	75.8	75.8	75.8	100.0	67.0			52.7	42.9
64	34.7	72.5	59.5	59.5	59.5	59.5	59.5	59.5	59.5	59.5	72.5	72.5			34.7	63.8
65	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	15.2
66	83.4	100.0	94.1	94.1	94.1	94.1	94.1	94.1	94.1	94.1	94.1	100.0			83.4	26.9
67	67.6	81.9	74.3	74.3	74.3	74.3	74.3	74.3	74.3	74.3	81.9	81.9			67.6	41.4
68	41.2	79.2	79.2	79.2	79.2	79.2	79.2	79.2	79.2	79.2	50.6	55.6			70.8	37.6
69	58.0	92.0	58.0	58.0	58.0	58.0	58.0	58.0	58.0	58.0	92.0	92.0			72.0	37.0
70	72.1	100.0	86.7	86.7	86.7	86.7	86.7	86.7	86.7	86.7	72.1	72.1		s00-----/32	100.0	13.0
71	42.0	76.0	54.0	54.0	54.0	54.0	54.0	54.0	54.0	54.0	71.7	71.7		s11--10-/8	46.0	51.7
72	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	11.9
73	69.6	100.0	74.6	74.6	74.6	74.6	74.6	74.6	74.6	74.6	69.6	69.6		s0------/64	100.0	13.5
74	51.7	65.5	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	65.5	56.2			51.7	50.8
75	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	21.4
76	42.7	85.8	79.5	79.5	79.5	79.5	79.5	79.5	79.5	79.5	73.8	73.8		s00-----/32	85.8	21.1
77	56.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	56.4	100.0			81.8	56.4
78	34.9	34.9	34.9	34.9	34.9	34.9	34.9	34.9	34.9	34.9	34.9	34.9			34.9	61.8
79	40.0	89.5	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0		s-1--1--/24	40.0	51.4
80	50.6	50.6	50.6	50.6	50.6	50.6	50.6	50.6	50.6	50.6	50.6	50.6			50.6	49.7
81	42.5	85.8	85.8	85.8	85.8	85.8	58.2	58.2	58.2	58.2	61.2	63.1			42.5	56.6
82	70.1	87.3	87.3	87.3	87.3	87.3	87.3	87.3	87.3	87.3	87.3	87.3			87.3	22.7
83	66.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	20.5
84	28.7	78.9	76.2	76.2	76.2	76.2	76.2	76.2	76.2	76.2	78.9	66.6			28.7	67.6
85	77.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	18.0
86	62.1	93.9	62.1	62.1	62.1	62.1	69.3	69.3	69.3	69.3	69.3	69.3		s0-----0/32	93.9	16.3
87	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	14.6
88	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	28.8
89	38.7	91.9	79.5	79.5	79.5	79.5	67.5	67.5	67.5	67.5	91.9	76.0			38.7	57.3
90	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6	96.6			96.6	10.1
91	91.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	15.8
92	65.7	65.7	65.7	65.7	65.7	65.7	65.7	65.7	65.7	65.7	65.7	65.7			65.7	39.0
93	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	19.7
94	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	13.1
95	47.7	63.4	47.7	47.7	47.7	47.7	47.7	47.7	47.7	47.7	63.4	63.4			52.9	44.4
96	59.3	100.0	59.3	59.3	59.3	59.3	59.3	59.3	59.3	59.3	59.3	59.3		s0------/64	100.0	20.7
97	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0			100.0	19.0
98	51.7	100.0	96.2	96.2	96.2	96.2	96.2	96.2	96.2	96.2	88.9	92.5		s00-----/32	100.0	15.3
99	48.8	85.7	85.7	85.7	85.7	85.7	85.7	85.7	85.7	85.7	75.4	85.7			66.0	39.4
AVE	71.7	91.0	84.6	84.6	84.6	84.6	84.5	84.5	84.4	84.4	84.3	84.3			81.8	28.3

KEY:
1: s1011111 2: s1001111 3: s1010111 4: s1000111 5: s1011110 6: s1001110 7: s1010110
8: s1000110 9: s1001100 10: s1010100 C:

Ranked document analysis

DOCUMENTS
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile.ranked

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[|]	[?]
00	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
01	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
02	4.5	120.5	108.5	108.5	108.5	108.5	96.5	96.5	66.5	108.5	108.5	96.5		s01--11-/8		
03	2.5	124.5	78.5	78.5	78.5	78.5	94.5	94.5	94.5	2.5	2.5	12.5				
04	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
05	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
06	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
07	2.5	126.5	28.5	28.5	28.5	28.5	13.5	13.5	16.5	8.5	8.5	2.5				
08	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
09	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
10	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
11	8.5	112.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	112.5	112.5	112.5				
12	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
13	4.5	120.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5	48.5		s-1-111-/8		
14	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
15	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
16	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
17	8.5	120.5	40.5	40.5	40.5	40.5	8.5	8.5	8.5	40.5	40.5	8.5				
18	16.5	126.5	44.5	44.5	44.5	44.5	44.5	44.5	94.5	44.5	44.5	44.5		s00-----/32		
19	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
20	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
21	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
22	32.5	120.5	104.5	104.5	104.5	104.5	120.5	120.5	120.5	104.5	104.5	120.5		s0------/64		
23	32.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	32.5	32.5	32.5				
24	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
25	22.5	120.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	62.5	62.5	22.5				
26	8.5	124.5	84.5	84.5	84.5	84.5	92.5	92.5	92.5	84.5	84.5	92.5		s10--0--/16		
27	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
28	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
29	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
30	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
31	4.5	120.5	36.5	36.5	36.5	36.5	4.5	4.5	4.5	36.5	36.5	4.5				
32	4.5	96.5	28.5	28.5	28.5	28.5	28.5	28.5	4.5	96.5	96.5	96.5				
33	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
34	32.5	120.5	72.5	72.5	72.5	72.5	120.5	120.5	120.5	72.5	72.5	120.5		s0------/64		
35	16.5	120.5	104.5	104.5	104.5	104.5	104.5	104.5	104.5	16.5	16.5	16.5				
36	1.0	120.5	3.5	7.5	3.5	7.5	5.0	6.0	2.0	42.5	42.5	25.5		s1011101/1		
37	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
38	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
39	32.5	104.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
40	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
41	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
42	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
43	32.5	120.5	88.5	88.5	88.5	88.5	120.5	120.5	120.5	88.5	88.5	120.5		s0------/64		
44	32.5	124.5	104.5	104.5	104.5	104.5	116.5	116.5	124.5	104.5	104.5	116.5		s0------/64		
45	48.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5		s-------/96		
46	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
47	4.5	96.5	16.5	16.5	16.5	16.5	28.5	28.5	4.5	16.5	16.5	28.5				
48	4.5	124.5	104.5	104.5	104.5	104.5	116.5	116.5	124.5	104.5	104.5	116.5		s1--100-/8		
49	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				

continued below

DOCUMENTS (continued)
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile.ranked

			OVERALL RANKED CONFIGURATIONS													
DOC	MIN	MAX	1	2	3	4	5	6	7	8	9	10	(C)	sTLSPCFU/Q	[|]	[?]
50	4.5	126.5	4.5	4.5	4.5	4.5	120.5	120.5	126.5	4.5	4.5	120.5				
51	16.5	96.5	16.5	16.5	16.5	16.5	16.5	16.5	16.5	16.5	16.5	16.5				
52	24.5	88.5	24.5	24.5	24.5	24.5	24.5	24.5	88.5	24.5	24.5	24.5				
53	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
54	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
55	4.5	124.5	4.5	4.5	4.5	4.5	40.5	40.5	40.5	4.5	4.5	40.5				
56	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
57	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
58	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
59	2.5	112.5	18.5	18.5	18.5	18.5	18.5	18.5	2.5	112.5	112.5	112.5				
60	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
61	2.5	112.5	6.5	6.5	14.5	14.5	76.5	76.5	52.5	6.5	6.5	76.5		s11-111-/4		
62	4.5	120.5	28.5	28.5	28.5	28.5	4.5	4.5	4.5	28.5	28.5	4.5				
63	4.5	96.5	36.5	36.5	36.5	36.5	60.5	60.5	4.5	36.5	36.5	60.5				
64	8.5	96.5	24.5	24.5	24.5	24.5	8.5	8.5	8.5	24.5	24.5	8.5				
65	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
66	4.5	96.5	20.5	20.5	20.5	20.5	4.5	4.5	20.5	20.5	20.5	4.5				
67	8.5	80.5	24.5	24.5	24.5	24.5	8.5	8.5	8.5	24.5	24.5	8.5				
68	4.5	127.5	4.5	4.5	4.5	4.5	86.5	86.5	106.5	4.5	4.5	86.5				
69	8.5	120.5	120.5	120.5	120.5	120.5	8.5	8.5	8.5	120.5	120.5	8.5				
70	16.5	124.5	116.5	116.5	116.5	116.5	124.5	124.5	124.5	116.5	116.5	124.5		s00-----/32		
71	4.5	124.5	76.5	76.5	76.5	76.5	12.5	12.5	12.5	76.5	76.5	12.5		s11--10-/8		
72	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
73	32.5	120.5	104.5	104.5	104.5	104.5	120.5	120.5	120.5	104.5	104.5	120.5		s0------/64		
74	4.5	96.5	36.5	36.5	36.5	36.5	36.5	36.5	4.5	36.5	36.5	36.5				
75	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
76	16.5	127.5	44.5	44.5	44.5	44.5	60.5	60.5	60.5	44.5	44.5	60.5		s00-----/32		
77	24.5	120.5	24.5	24.5	24.5	24.5	24.5	24.5	120.5	24.5	24.5	24.5				
78	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
79	12.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5		s-1--1--/24		
80	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
81	4.5	112.5	4.5	4.5	4.5	4.5	26.5	26.5	10.5	58.5	58.5	46.5				
82	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
83	32.5	120.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
84	3.5	112.5	10.5	10.5	10.5	10.5	24.5	24.5	3.5	10.5	10.5	24.5				
85	32.5	124.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
86	16.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	120.5	104.5	104.5	104.5		s0-----0/32		
87	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
88	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
89	2.5	120.5	12.5	12.5	12.5	12.5	16.5	16.5	2.5	45.5	45.5	16.5				
90	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
91	32.5	96.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5	32.5				
92	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
93	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
94	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
95	4.5	120.5	120.5	120.5	120.5	120.5	4.5	4.5	4.5	120.5	120.5	4.5				
96	32.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5	112.5		s0------/64		
97	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5	64.5				
98	16.5	124.5	44.5	44.5	44.5	44.5	58.5	58.5	62.5	44.5	44.5	58.5		s00-----/32		
99	6.5	124.5	6.5	6.5	6.5	6.5	6.5	6.5	15.5	6.5	6.5	6.5				
AVE	-	-	55.6	55.6	55.6	55.7	56.8	56.8	57.1	57.1	57.1	57.1			-	-

KEY:
1: s1011111 2: s1001111 3: s1010111 4: s1000111 5: s1010101 6: s1000101 7: s1001101
8: s1011110 9: s1001110 10: s1000100 C:

Configurations

Numeric configuration analysis

CONFIGURATIONS
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile

NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
1	s1011111	34.9	100.0	65.1	84.6	0	0	0	1	6	12	5	10	11	7	48	19.27
2	s1001111	34.9	100.0	65.1	84.6	0	0	0	1	6	12	5	11	10	7	48	19.27
3	s1010111	34.9	100.0	65.1	84.6	0	0	0	1	6	12	5	10	11	7	48	19.29
4	s1000111	34.9	100.0	65.1	84.6	0	0	0	1	6	12	5	11	10	7	48	19.29
5	s1011110	29.9	100.0	70.1	84.5	0	0	1	2	4	12	4	12	10	7	48	19.56
6	s1001110	29.9	100.0	70.1	84.5	0	0	1	2	4	12	4	12	10	7	48	19.56
7	s1010110	29.9	100.0	70.1	84.4	0	0	1	2	4	12	4	12	10	7	48	19.58
8	s1000110	29.9	100.0	70.1	84.4	0	0	1	2	4	12	4	12	10	7	48	19.58
9	s1001100	29.9	100.0	70.1	84.3	0	0	1	2	5	6	12	11	7	8	48	19.49
10	s1010100	29.9	100.0	70.1	84.3	0	0	1	2	4	7	12	10	8	7	49	19.49
11	s1000100	29.9	100.0	70.1	84.3	0	0	1	2	4	7	12	10	8	7	49	19.49
12	s1001101	34.9	100.0	65.1	84.2	0	0	0	1	5	9	13	9	8	8	47	18.80
13	s1011100	29.9	100.0	70.1	84.1	0	0	1	2	5	7	12	10	7	8	48	19.67
14	s1011101	34.9	100.0	65.1	84.1	0	0	0	1	5	10	13	8	8	8	47	18.99
15	s1010101	34.9	100.0	65.1	83.8	0	0	0	1	6	9	12	9	8	7	48	19.40
16	s1000101	34.9	100.0	65.1	83.8	0	0	0	1	6	9	12	9	8	7	48	19.40
17	s1011011	34.8	100.0	65.2	83.0	0	0	0	2	8	11	6	7	11	11	44	20.82
18	s1001011	34.8	100.0	65.2	83.0	0	0	0	2	8	11	6	7	11	11	44	20.83
19	s1010011	34.8	100.0	65.2	83.0	0	0	0	2	9	10	6	7	11	11	44	20.83
20	s1000011	34.8	100.0	65.2	83.0	0	0	0	2	9	10	6	7	11	11	44	20.84
21	s1001001	34.9	100.0	65.1	83.0	0	0	0	1	9	11	8	5	13	11	42	20.26
22	s1010001	34.8	100.0	65.2	83.0	0	0	0	3	7	10	8	6	11	11	44	21.09
23	s1000001	34.8	100.0	65.2	83.0	0	0	0	3	7	10	8	6	11	11	44	21.10
24	s1011001	34.9	100.0	65.1	83.0	0	0	0	1	9	11	8	5	13	11	42	20.29
25	s1010000	29.9	100.0	70.1	82.9	0	0	1	4	5	11	8	5	11	9	46	21.53
26	s1000000	29.9	100.0	70.1	82.9	0	0	1	4	5	11	8	5	11	9	46	21.53
27	s1011010	29.9	100.0	70.1	82.9	0	0	1	3	6	12	7	5	11	9	46	21.41
28	s1001010	29.9	100.0	70.1	82.9	0	0	1	3	6	12	7	5	11	9	46	21.41
29	s1000010	29.9	100.0	70.1	82.9	0	0	1	3	7	11	7	5	11	9	46	21.43
30	s1010010	29.9	100.0	70.1	82.9	0	0	1	3	7	11	7	5	11	9	46	21.43
31	s0011101	28.7	100.0	71.3	82.8	0	0	2	2	6	11	8	5	11	5	50	22.04
32	s0001101	28.7	100.0	71.3	82.8	0	0	2	2	6	11	8	5	11	5	50	22.04
33	s0000111	28.7	100.0	71.3	82.8	0	0	2	2	9	8	6	7	11	5	50	22.27
34	s0001111	28.7	100.0	71.3	82.8	0	0	2	2	9	8	6	7	11	5	50	22.27
35	s0011111	28.7	100.0	71.3	82.8	0	0	2	2	9	8	6	7	11	5	50	22.27
36	s0010111	28.7	100.0	71.3	82.8	0	0	2	2	9	8	6	7	11	5	50	22.27
37	s1001000	29.9	100.0	70.1	82.7	0	0	1	3	5	13	9	3	13	9	44	21.09
38	s1011000	29.9	100.0	70.1	82.7	0	0	1	3	5	13	9	3	13	9	44	21.12
39	s0000101	28.7	100.0	71.3	82.1	0	0	2	2	10	10	6	4	11	5	50	22.84
40	s0010101	28.7	100.0	71.3	82.1	0	0	2	2	10	10	6	4	11	5	50	22.84
41	s0001001	28.7	100.0	71.3	82.1	0	0	3	2	9	9	6	4	12	7	48	23.01
42	s0011001	28.7	100.0	71.3	82.1	0	0	3	2	9	9	6	4	12	7	48	23.01
43	s0001011	28.7	100.0	71.3	82.0	0	0	2	3	9	9	6	4	12	7	48	23.01
44	s0000011	28.7	100.0	71.3	82.0	0	0	2	3	9	9	6	4	12	7	48	23.01
45	s0010011	28.7	100.0	71.3	82.0	0	0	2	3	9	9	6	4	12	7	48	23.01
46	s0011011	28.7	100.0	71.3	82.0	0	0	2	3	9	9	6	4	12	7	48	23.01
47	s0000001	25.6	100.0	74.4	81.8	0	0	3	4	8	8	6	4	12	7	48	23.43
48	s0010001	25.6	100.0	74.4	81.8	0	0	3	4	8	8	6	4	12	7	48	23.43
49	s0001000	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
50	s0001010	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92

continued below

CONFIGURATIONS (continued)																	
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
51	s0000100	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
52	s0011110	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
53	s0000110	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
54	s0011010	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
55	s0001100	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
56	s0011000	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
57	s0011100	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
58	s0010110	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
59	s0000010	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
60	s0010100	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
61	s0001110	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
62	s0010010	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
63	s0000000	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
64	s0010000	19.3	100.0	80.7	81.8	0	1	3	4	7	7	7	4	11	6	50	23.92
65	s1110001	34.8	100.0	65.2	81.3	0	0	0	2	12	7	9	6	15	13	36	20.60
66	s1100001	34.8	100.0	65.2	81.3	0	0	0	2	12	8	8	6	15	13	36	20.64
67	s1110000	29.9	100.0	70.1	81.2	0	0	1	1	11	9	9	5	15	12	37	21.02
68	s1100000	29.9	100.0	70.1	81.2	0	0	1	1	11	10	8	5	15	12	37	21.06
69	s1101001	34.9	100.0	65.1	81.0	0	0	0	1	9	13	9	5	16	13	34	19.82
70	s1111001	34.9	100.0	65.1	81.0	0	0	0	1	9	13	9	5	16	13	34	19.82
71	s1110100	29.9	100.0	70.1	80.9	0	0	1	2	9	8	12	12	7	9	40	20.51
72	s1111111	34.9	100.0	65.1	80.9	0	0	0	1	9	13	8	11	11	8	39	20.34
73	s1101111	34.9	100.0	65.1	80.8	0	0	0	1	9	13	7	12	11	8	39	20.37
74	s1100100	29.9	100.0	70.1	80.8	0	0	1	2	9	8	12	12	7	9	40	20.60
75	s1110111	34.9	100.0	65.1	80.8	0	0	0	1	9	13	8	11	11	9	38	20.31
76	s1101110	29.9	100.0	70.1	80.8	0	0	1	1	7	15	7	11	10	9	39	20.54
77	s1111110	29.9	100.0	70.1	80.8	0	0	1	1	7	15	8	10	10	9	39	20.54
78	s1100111	34.9	100.0	65.1	80.8	0	0	0	1	9	13	7	12	11	9	38	20.34
79	s1100110	29.9	100.0	70.1	80.7	0	0	1	1	7	15	7	11	10	10	38	20.52
80	s1110110	29.9	100.0	70.1	80.7	0	0	1	1	7	15	8	10	10	10	38	20.52
81	s1111000	29.9	100.0	70.1	80.7	0	0	1	1	8	15	9	3	16	12	35	20.61
82	s1101000	29.9	100.0	70.1	80.7	0	0	1	1	8	15	9	3	16	12	35	20.62
83	s1111011	34.8	100.0	65.2	80.6	0	0	0	2	11	12	6	8	13	12	36	20.72
84	s1110011	34.8	100.0	65.2	80.6	0	0	0	2	11	12	6	8	13	12	36	20.73
85	s1101100	29.9	100.0	70.1	80.6	0	0	1	2	11	6	12	12	6	11	39	20.92
86	s1101011	34.8	100.0	65.2	80.6	0	0	0	2	11	13	5	8	13	12	36	20.75
87	s1100011	34.8	100.0	65.2	80.6	0	0	0	2	11	13	5	8	13	12	36	20.77
88	s1111010	29.9	100.0	70.1	80.5	0	0	1	1	10	15	6	6	13	11	37	21.09
89	s1110010	29.9	100.0	70.1	80.5	0	0	1	1	10	15	6	6	13	11	37	21.11
90	s1101010	29.9	100.0	70.1	80.5	0	0	1	1	10	16	5	6	13	11	37	21.13
91	s1100010	29.9	100.0	70.1	80.5	0	0	1	1	10	16	5	6	13	11	37	21.15
92	s1111100	29.9	100.0	70.1	80.4	0	0	1	2	12	6	11	13	5	11	39	21.10
93	s1101101	34.9	100.0	65.1	80.4	0	0	0	2	10	8	14	11	7	10	38	20.35
94	s1110101	34.9	100.0	65.1	80.3	0	0	0	2	10	9	13	12	7	8	39	20.39
95	s1111101	34.9	100.0	65.1	80.3	0	0	0	2	11	8	13	12	6	10	38	20.51
96	s1100101	34.9	100.0	65.1	80.2	0	0	0	2	10	9	13	12	7	8	39	20.51
97	s0110001	28.8	100.0	71.2	80.1	0	0	2	4	11	7	7	4	16	12	37	22.71
98	s0100001	28.8	100.0	71.2	80.1	0	0	2	4	11	7	7	4	16	12	37	22.74
99	s0110000	21.2	100.0	78.8	80.0	0	0	3	3	11	7	8	3	15	12	38	23.25
100	s0111111	28.8	100.0	71.2	80.0	0	0	1	3	11	8	8	9	13	8	39	22.08
continued below																	

CONFIGURATIONS (continued) /hame/jeremyc/phd/eval/sys/scores/TEST/scorefile																	
NUM	CONFIG	MIN	MAX	VAR	X	DISTRIBUTION											STDEV
	sTLSPCFU					0	10	20	30	40	50	60	70	80	90	00	
101	s0100000	21.2	100.0	78.8	80.0	0	0	3	3	11	7	8	3	15	12	38	23.28
102	s0111001	28.8	100.0	71.2	80.0	0	0	1	4	10	10	7	4	16	11	37	22.20
103	s0101111	28.8	100.0	71.2	79.9	0	0	1	3	12	7	8	9	13	8	39	22.15
104	s0110111	28.8	100.0	71.2	79.9	0	0	1	3	11	8	8	9	13	9	38	22.03
105	s0101001	28.8	100.0	71.2	79.9	0	0	1	4	10	10	7	4	16	11	37	22.23
106	s0100111	28.8	100.0	71.2	79.9	0	0	1	3	12	7	8	9	13	9	38	22.11
107	s0111011	28.8	100.0	71.2	79.7	0	0	1	4	10	11	6	6	14	11	37	22.47
108	s0110011	28.8	100.0	71.2	79.7	0	0	1	4	10	11	6	6	14	11	37	22.47
109	s0100011	28.8	100.0	71.2	79.7	0	0	1	4	10	11	6	6	14	11	37	22.50
110	s0101011	28.8	100.0	71.2	79.7	0	0	1	4	10	11	6	6	14	11	37	22.50
111	s0111000	23.7	100.0	76.3	79.6	0	0	3	3	11	8	8	3	15	11	38	23.14
112	s0110010	28.8	100.0	71.2	79.6	0	0	2	3	13	8	7	5	13	11	38	23.00
113	s0111010	28.8	100.0	71.2	79.6	0	0	2	3	13	8	7	5	13	11	38	23.00
114	s0101000	23.7	100.0	76.3	79.6	0	0	3	3	11	8	8	3	15	11	38	23.17
115	s0111110	26.9	100.0	73.1	79.5	0	0	3	3	11	8	6	8	12	10	39	23.11
116	s0100010	28.8	100.0	71.2	79.5	0	0	2	3	13	8	7	5	13	11	38	23.03
117	s0101010	28.8	100.0	71.2	79.5	0	0	2	3	13	8	7	5	13	11	38	23.03
118	s0101110	26.9	100.0	73.1	79.5	0	0	3	3	11	8	6	8	12	10	39	23.15
119	s0110110	26.9	100.0	73.1	79.5	0	0	3	3	11	8	6	8	12	11	38	23.06
120	s0111101	28.8	100.0	71.2	79.4	0	0	1	3	9	12	10	9	10	8	38	21.58
121	s0100110	26.9	100.0	73.1	79.4	0	0	3	3	11	8	6	8	12	11	38	23.10
122	s0101101	28.8	100.0	71.2	79.4	0	0	1	3	9	12	10	9	10	8	38	21.64
123	s0111100	26.7	100.0	73.3	79.1	0	0	3	3	11	10	7	8	9	10	39	23.06
124	s0110100	22.4	100.0	77.6	79.1	0	0	3	3	11	9	9	7	9	10	39	23.14
125	s0101100	26.7	100.0	73.3	79.0	0	0	3	3	11	10	7	8	9	10	39	23.10
126	s0100100	22.4	100.0	77.6	79.0	0	0	3	3	12	8	9	7	9	10	39	23.22
127	s0110101	28.8	100.0	71.2	78.9	0	0	1	3	11	11	11	7	10	8	38	22.11
128	s0100101	28.8	100.0	71.2	78.8	0	0	1	3	12	10	11	7	10	8	38	22.21
	AVE	28.9	100.0	71.1	81.5	0.0	0.1	1.4	2.5	8.6	9.8	7.7	6.8	11.4	8.9	42.8	21.71

Ranked configuration analysis

CONFIGURATIONS
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
1	s1011111	3	120	117	55.6	61	12	11	7	3	1	1	2	1	1	0	30.71
2	s1001111	4	120	116	55.6	61	12	10	7	3	2	1	2	1	1	0	30.64
3	s1010111	3	120	117	55.6	61	11	11	8	3	1	1	2	1	1	0	30.59
4	s1000111	4	120	116	55.7	61	11	10	8	3	2	1	2	1	1	0	30.52
5	s1010101	4	124	120	56.8	65	3	10	8	7	1	0	2	1	1	4	32.99
6	s1000101	4	124	120	56.8	65	3	10	7	8	1	0	2	1	1	4	32.98
7	s1001101	2	126	124	57.1	68	7	6	6	4	2	1	2	2	0	3	35.84
8	s1011110	2	120	118	57.1	61	14	8	5	3	1	2	1	2	1	2	29.71
9	s1001110	2	120	118	57.1	61	14	8	5	3	1	2	1	2	1	2	29.71
10	s1000100	2	124	122	57.1	67	3	8	5	7	1	2	3	1	1	3	32.58
11	s1010100	2	124	122	57.1	67	3	8	5	7	1	2	3	1	1	3	32.58
12	s1010110	2	120	118	57.2	61	13	8	6	3	1	2	1	2	1	2	29.58
13	s1000110	2	120	118	57.2	61	13	8	6	3	1	2	1	2	1	2	29.58
14	s1010001	8	120	112	57.5	55	17	8	3	4	4	2	0	1	1	6	20.39
15	s1000001	8	120	112	57.5	55	17	8	3	4	4	2	0	1	1	6	20.40
16	s1001001	4	120	116	57.5	54	18	8	2	6	4	0	1	2	0	6	22.96
17	s1001100	2	126	124	57.8	68	5	6	5	4	2	2	5	2	0	1	34.93
18	s1011001	4	120	116	57.9	54	18	8	1	6	5	0	1	2	0	6	22.82
19	s1011101	1	126	125	58.0	69	5	6	6	4	2	1	2	2	1	3	35.99
20	s1010000	8	120	112	58.0	57	15	9	2	4	2	2	1	1	3	5	23.08
21	s1000000	8	120	112	58.0	57	15	9	2	4	2	2	1	1	3	5	23.08
22	s1001000	4	120	116	58.7	56	15	9	2	5	2	1	2	2	2	5	24.90
23	s1011100	2	126	124	58.7	68	4	6	5	4	2	2	5	2	1	1	35.05
24	s1011011	8	120	112	59.0	55	14	10	4	6	3	1	1	0	0	7	22.23
25	s1001011	8	120	112	59.1	55	14	10	4	6	3	1	1	0	0	7	22.23
26	s1010011	8	120	112	59.1	55	14	10	4	6	2	2	1	0	0	7	22.15
27	s1011000	4	120	116	59.1	56	15	9	1	5	3	1	2	2	2	5	24.78
28	s1000011	8	120	112	59.1	55	14	10	4	6	2	2	1	0	0	7	22.14
29	s1001010	6	120	114	59.8	57	13	10	2	6	2	1	2	0	2	6	24.74
30	s1011010	6	120	114	59.8	57	13	10	2	6	2	1	2	0	2	6	24.74
31	s1010010	6	120	114	59.8	57	13	10	2	6	1	2	2	0	2	6	24.66
32	s1000010	6	120	114	59.8	57	13	10	2	6	1	2	2	0	2	6	24.66
33	s0000111	8	118	110	60.0	63	4	12	4	6	1	2	1	0	1	8	26.54
34	s0001111	8	118	110	60.0	63	4	12	4	6	1	2	1	0	1	8	26.54
35	s0010111	8	118	110	60.0	63	4	12	4	6	1	2	1	0	1	8	26.54
36	s0011111	8	118	110	60.0	63	4	12	4	6	1	2	1	0	1	8	26.54
37	s0001101	2	126	124	60.2	65	4	11	3	6	1	2	0	0	1	9	26.69
38	s0011101	2	126	124	60.2	65	4	11	3	6	1	2	0	0	1	9	26.69
39	s0010101	8	126	118	61.6	63	4	11	4	6	1	2	0	1	1	9	25.84
40	s0000101	8	126	118	61.6	63	4	11	4	6	1	2	0	1	1	9	25.84
41	s0011001	12	112	100	62.2	61	8	9	2	6	3	3	0	0	1	10	24.17
42	s0001001	12	112	100	62.2	61	8	9	2	6	3	3	0	0	1	10	24.17
43	s0010011	16	116	100	63.0	61	7	9	3	6	2	3	0	1	1	10	23.71
44	s0011011	16	116	100	63.0	61	7	9	3	6	2	3	0	1	1	10	23.71
45	s0001011	16	116	100	63.0	61	7	9	3	6	2	3	0	1	1	10	23.71
46	s0000011	16	116	100	63.0	61	7	9	3	6	2	3	0	1	1	10	23.71
47	s0010001	16	112	96	63.0	61	7	9	3	6	2	4	0	0	1	9	23.92
48	s0000001	16	112	96	63.0	61	7	9	3	6	2	4	0	0	1	9	23.92
49	s0000100	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
50	s0000110	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02

continued below

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
51	s0001000	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
52	s0011100	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
53	s0001010	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
54	s0011010	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
55	s0001100	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
56	s0011000	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
57	s0000010	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
58	s0010110	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
59	s0001110	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
60	s0010100	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
61	s0011110	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
62	s0010010	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
63	s0010000	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
64	s0000000	16	120	104	63.1	64	4	8	3	7	2	3	1	0	1	9	27.02
65	s1110001	20	120	100	65.4	44	21	10	7	3	4	1	1	2	2	6	18.45
66	s1101111	2	124	122	65.5	49	12	16	6	3	1	5	1	2	1	5	32.60
67	s1100001	20	120	100	65.5	44	21	10	7	2	5	1	1	3	1	6	18.60
68	s1100111	4	124	120	65.6	47	13	17	6	3	1	5	1	2	1	5	32.31
69	s1110000	20	120	100	65.6	45	21	9	5	4	3	3	1	2	2	7	18.74
70	s1100000	20	120	100	65.7	45	21	9	5	3	4	3	1	3	1	7	18.86
71	s1111001	4	120	116	65.8	43	19	15	6	4	2	1	2	1	2	6	23.34
72	s1101001	4	120	116	65.8	43	19	15	6	3	3	2	1	0	2	8	23.21
73	s1111111	2	124	122	65.8	49	11	16	7	3	1	4	4	1	0	5	32.33
74	s1101110	2	120	118	65.8	48	13	15	7	2	1	6	1	2	1	5	29.57
75	s1100110	10	120	110	66.0	46	14	16	7	2	1	6	1	2	1	5	29.25
76	s1110111	4	124	120	66.0	47	12	17	7	3	1	4	4	1	0	5	32.04
77	s1110100	4	124	120	66.2	52	11	7	12	6	0	4	1	2	0	6	34.23
78	s1100100	4	127	123	66.3	52	10	8	12	5	0	5	1	3	0	4	34.25
79	s1111110	2	120	118	66.3	48	12	15	8	1	2	5	3	2	0	5	29.06
80	s1110110	10	120	110	66.5	46	13	16	8	1	2	5	3	2	0	5	28.73
81	s1101000	4	120	116	66.8	44	18	13	5	4	3	4	1	0	2	9	22.18
82	s1111000	4	120	116	66.8	44	18	13	5	5	2	3	2	1	2	7	22.31
83	s1110101	4	124	120	66.9	51	9	9	13	8	0	1	1	2	1	6	35.76
84	s1111011	18	124	106	67.0	44	18	11	6	8	1	1	2	2	0	8	21.51
85	s1101011	18	124	106	67.0	44	18	11	7	6	2	1	2	2	0	8	21.63
86	s1110011	20	124	104	67.1	44	18	11	6	7	2	1	2	2	0	8	21.38
87	s1100011	20	124	104	67.1	44	18	11	7	5	3	1	2	2	0	8	21.50
88	s1111010	18	124	106	67.1	45	18	10	4	8	1	3	2	2	1	8	21.32
89	s1100101	4	127	123	67.1	51	8	10	12	7	0	3	1	3	0	6	35.66
90	s1101010	18	124	106	67.2	45	18	10	5	6	2	3	2	2	1	8	21.45
91	s1110010	20	124	104	67.2	45	18	10	4	7	2	3	2	2	1	8	21.19
92	s1101100	4	126	122	67.2	53	11	8	8	5	4	3	2	2	0	4	37.14
93	s1100010	20	124	104	67.2	45	18	10	5	5	3	3	2	2	1	8	21.32
94	s1101101	4	126	122	67.4	53	11	9	7	6	4	2	1	3	0	4	39.25
95	s1111100	3	127	124	67.8	54	12	5	8	4	5	3	2	2	0	5	37.55
96	s1111101	3	127	124	67.9	54	12	6	8	5	5	2	1	2	0	5	39.77
97	s0110001	32	120	88	70.3	47	16	7	10	7	1	2	2	0	2	6	17.82
98	s0100001	32	120	88	70.4	47	16	7	9	8	1	2	2	0	2	6	17.85
99	s0111001	10	120	110	70.4	47	13	10	9	7	2	2	2	0	2	7	20.24
100	s0110000	16	120	104	70.5	49	14	7	8	8	1	4	2	0	1	6	19.19

continued below

CONFIGURATIONS (continued)
/hame/jeremyc/phd/eval/sys/scores/TEST/scorefile.ranked

NUM	CONFIG	TOP	BTM	VAR	$\bar{X}$	DISTRIBUTION											STDEV
	sTLSPCFU					1	2	3	4	5	6	7	8	9	10	++	
101	s0101001	10	120	110	70.5	47	13	10	8	8	2	2	2	0	2	7	20.25
102	s0111101	4	124	120	70.6	50	9	11	7	8	2	2	1	1	3	7	29.62
103	s0100000	16	120	104	70.6	49	14	7	7	9	1	4	2	0	1	6	19.21
104	s0101101	4	124	120	70.7	50	9	11	7	8	2	2	1	0	3	9	29.75
105	s0111111	4	124	120	70.9	51	5	16	7	5	1	4	0	0	2	9	27.68
106	s0110111	4	124	120	71.0	50	6	16	7	5	1	4	0	0	2	9	27.50
107	s0101111	4	124	120	71.1	51	5	16	7	5	1	3	1	0	2	9	27.77
108	s0111100	6	124	118	71.2	51	9	8	7	8	2	4	1	1	3	7	27.79
109	s0100111	4	124	120	71.2	50	6	16	7	5	1	3	1	0	2	9	27.59
110	s0101100	6	124	118	71.2	51	9	8	7	8	2	4	1	0	3	9	27.90
111	s0110011	32	121	89	71.4	47	15	8	8	7	1	1	3	0	2	10	19.12
112	s0111011	32	121	89	71.4	47	15	8	8	7	1	1	3	0	2	10	19.12
113	s0110100	16	124	108	71.4	50	9	8	9	7	1	4	1	1	3	8	25.62
114	s0111000	16	120	104	71.5	49	10	9	8	9	2	4	2	0	1	6	20.23
115	s0101011	32	123	91	71.5	47	15	8	7	8	1	1	3	0	2	10	19.22
116	s0100011	32	123	91	71.5	47	15	8	7	8	1	1	3	0	2	10	19.22
117	s0100100	16	124	108	71.5	50	9	8	9	7	1	4	1	0	3	9	25.81
118	s0111010	16	120	104	71.6	49	13	8	6	8	1	3	3	0	2	8	19.81
119	s0110010	16	120	104	71.6	49	13	8	6	8	1	3	3	0	2	8	19.81
120	s0101000	16	120	104	71.6	49	10	9	7	10	2	4	2	0	1	6	20.25
121	s0101010	16	120	104	71.7	49	13	8	5	9	1	3	3	0	2	8	19.82
122	s0100010	16	120	104	71.7	49	13	8	5	9	1	3	3	0	2	8	19.82
123	s0110101	16	124	108	71.8	48	8	11	9	8	1	2	1	2	3	8	25.89
124	s0111110	4	124	120	71.9	52	6	13	7	4	1	6	0	0	2	9	25.56
125	s0101110	4	124	120	71.9	52	6	13	7	4	1	5	1	0	2	9	25.63
126	s0110110	4	124	120	71.9	51	7	13	7	4	1	6	0	0	2	9	25.36
127	s0100101	16	124	108	72.0	48	8	11	9	8	1	2	1	1	3	9	26.10
128	s0100110	4	124	120	72.0	51	7	13	7	4	1	5	1	0	2	9	25.43
	AVE	11.0	121.9	110.9	64.5	54	10	9	5	5	1	2	1	0	1	6	26.07

Components

Numeric (left) and ranked (right) component analysis

sT------	1.361840	sT------	4.652500
s-L-----	-2.691750	s-L-----	-8.802188
s--S----	0.008032	s--S----	-0.041562
s---P---	0.006010	s---P---	-0.091875
s----C--	0.291661	s----C--	0.695625
s-----F-	0.100181	s-----F-	-0.136563
s------U	0.196969	s------U	0.750156

Appendix D

Correlation set

D.1 Instructions to coders

Event distinction experiment

This is the starting page for my short experiment. Please read the instructions below before continuing.

Instructions

You'll be presented with a brief news article relating to some aspect of Latin American terrorist incidents. The article will be split up into boxes, with the big boxes on the left representing single paragraphs.
News articles of this nature typically describe pretty gruesome events. Sometimes several different incidents are described in the same news article, perhaps to provide some sort of background information for the reader, or simply because it's been a busy day.
Your task in this experiment is a simple one: you have to split the text up into different events. Accompanying each "line" of text is a series of buttons. By clicking on these buttons, you can specify how these lines of text relate to events. For example, consider the following short text. (The first couple of lines of each text usually contain header information.)

S	C	Text	Grid									
0	1	Bogota,	□									
0	2	4 Feb 90 (acan efe).	□									
1	1	A bomb exploded yesterday in downtown Aracataca.		□								
2	1	Police in Bogota say that the JPF were responsible for the attack.		□								
3	1	Several buildings were damaged in the blast.		□								
4	1	In a similar incident ten days ago,			□							
4	2	the JPF attacked an army installation in the town of Rivera.			□							
5	1	The guards shot two terrorists.			□							
6	1	JPF guerillas are known to frequent the town.			□							
7	1	Saturday's attack occurred at around 1725 local time,		□								
7	2	and may have been timed to disrupt rush hour traffic.		□								

In this example, the fragment starting "In a similar incident ten days ago..." and ending "... the town" has been designated as referring to a different event from the rest of the text on either side. Consequently, the buttons have been pushed to show this. If the reader had thought that this text only described one event (or no events!), all the lit buttons would have been aligned in the same column. If, on the other hand, they had believed that every line referred to a new event (unlikely!), the lit buttons would have gone down to the right diagonally. (In fact, there's not enough buttons to do this, but you get the idea.)
So, your task is to segment the texts by clicking the buttons that accompany each line. The texts I've included vary a lot in form and number of events; not all will have more than one event in them. If you think there's no more than one event in the text, just hit the "submit" button at the bottom of the page.
There are eight separate texts to segment. Please read each one carefully - it's a good idea to read through it once before starting to segment the text. Also, please use the buttons from left to right in each text, starting back at the left-hand side for each new text. (Treat each text in isolation, don't look for similar events between texts; you can completely forget about a text once you've finished it! Also, ignore the order in which things happen - I'm only interested in whether it's a different event). Only use a new button if you think that a new event is being described. If you need to make any changes, just submit the text again.
After the final text, there's a short description of what all this is in aid of.
Thanks for taking part!
Jeremy Crowe

<u>Start the experiment</u>

D.2 Corpus

Document 1		guatemala city , 4 feb 90 (acan efe).
1	1	the guatemala army denied today that guerrillas attacked the santo tomas presidential farm ,
	2	located on the pacific side ,
	3	where president cerezo has been staying since 2 february.
2	1	a report published by the cerigua news agency mouthpiece of the guatemalan national revolutionary unity (urng) whose main offices are in mexico , says that a guerrilla column attacked the farm 2 days ago.
3	1	however , armed forces spokesman colonel luis arturo isaacs said that the attack ,
	2	which resulted in the death of a civilian
	3	who was passing by at the time of the skirmish ,
	4	was not against the farm ,
	5	and that president cerezo is safe and sound.
4	1	he added that on 3 february president cerezo met with the diplomatic corps accredited in guatemala.
5	1	the government also issued a communique describing the rebel report as false and incorrect and stressing that the president was never in danger.
6	1	col isaacs said that the guerrillas attacked the la eminencia farm located near the santo tomas farm ,
	2	where they burned the facilities and stole food.
7	1	a military patrol clashed with a rebel column and inflicted three casualties ,
	2	which were taken away by the guerrillas
	3	who fled to the mountains ,
	4	isaacs noted.
8	1	he also reported that guerrillas killed a peasant in the city of flores ,
	2	in the northern el peten department ,
	3	and burned a tank truck.

Document 2		bogota , 18 aug 89 (inravision television cadena 1).
1	1	the extraditables today claimed responsibility for the murder of antioquia police commander colonel waldemar franklin quintero ,
	2	which occurred this morning in medellin.
2	1	the criminal attack on the officer occured a scant 1 month after the murder of antioquia department governor antonio roldan betancur.
3	1	a policeman was killed and another policeman wounded during the terrorist attack.
4	1	col quintero did not have a police escort by his own behest.
5	1	col quintero was directing operations against drug trafficking and consequently had been the object of death threats.
6	1	here is a report by gladys vargas from medellin.
7	1	(begin recording) (vargas) this morning antioquia police commander col waldemar franklin quintero left his residence to begin his daily tasks at antioquia police headquarters.
8	1	as the vehicle drove through the america neighborhood ,
	2	seven blocks from
	3	where antioquia governor antonio roldan was killed ,
	4	the vehicle that col quintero was in a white nissan patrol vehicle with license plates me 7847 drew to a stop in front of a red light at the carrera 80 crossing.
9	1	several individuals in a gray mazda 626 vehicle ,
	2	with license plates mi 1630 ,
	3	took advantage of this circumstance to fire volleys of submachinegun fire against col quintero ,
	4	who died instantly.
10	1	(passage omitted)
11	1	col quintero did not have a police escort.
12	1	just 8 days ago he decided to dismiss the escort ,
	2	because he did not want anyone else to be killed
	3	if there were an attack on him.
13	1	only col quintero and a police driver were in the vehicle.
14	1	the judge of the 76 th district court , who directed the removal of the body , said that several types of weapons were used in the attack and more than 100 shots were fired.
15	1	(passage omitted) (end recording)

Document 3		bogota , 27 apr 90 (inravision television cadena 1).
1	1	(colombian president virgilio barco) (text) (continued) the media can make a significant contribution toward peace and democracy
	2	if ,
	3	on their own initiative ,
	4	they launch a great civic campaign to vigorously stimulate the citizens to participate in the coming elections.
2	1	fellow citizens , i know that you are asking yourselves
	2	what can be done under these circumstances.
3	1	above all ,
	2	do not play up to terrorism.
4	1	the interests of the fatherland are above all individual or group interests ,
	2	as is the defense of our democracy ,
	3	which we have all been building together for more than a century.
5	1	this cannot be the time of opportunism.
6	1	this is the hour of solidarity with the country ,
	2	its people ,
	3	and its institutions.
7	1	the effectiveness of these measures and the actions of the public force depend ,
	2	first of all ,
	3	on the solidarity of the citizens.
8	1	the results will be as encouraging and as favorable as the cooperation of the citizens.
9	1	to regain tranquility is not the exclusive responsibility of the armed forces or the state.
10	1	it is a task in
	2	which each of us has something to contribute.
11	1	colombia has successfully overcome even more difficult moments.
12	1	it has faced decisively and emerged victoriously from even more uncertain situations.
13	1	that past has created solid institutions.
14	1	for this reason ,
	2	i am sure that democracy will triumph
	3	and that those
	4	who commit violence ,
	5	regardless of
	6	who they may be ,
	7	will be punished.
15	1	i am sure that we will continue to progress along the path of peace ,
	2	tolerance ,
	3	and respect for life and human dignity.
16	1	good night.

Document 4		san salvador , 28 mar 89.	
1	1	military authorities have reported that two peasants were killed	
	2	when they set off a mine in morazan department.	
2	1	the incident took place in the vicinity of perquin ,	
	2	when ernesto and rosa fernandez set off an explosive device presumably abandoned by the fmln.	
3	1	it was also reported that four treasury police officers were wounded last night	
	2	when urban commandos attacked them in soyapango.	
4	1	the incident took place on army boulevard while the officers were on their way to start guard duty at a shoe factory.	
5	1	in another action ,	
	2	six insurgents were wounded during a clash with 1 st brigade soldiers in san pedro perulapan jurisdiction ,	
	3	cuscatlan department.	

Document 5		san salvador , 19 mar 90 (acan efe).	
1	1	a bomb went off this morning near a power tower in san salvador leaving a large part of the city without energy ,	
	2	but no casualties have been reported.	
2	1	according to unofficial sources ,	
	2	the bomb allegedly detonated by urban guerrilla commandos blew up a power tower in the northwestern part of san salvador at 650 (1250 gmt).	
3	1	on friday ,	
	2	16 march ,	
	3	the farabundo marti national liberation front (fmln) decreed a suspension of its attacks on civilian government officials and of its sabotage operations against means of transportation ,	
	4	businesses ,	
	5	and telecommunications.	
4	1	that decision ,	
	2	aimed at creating the appropriate conditions for a new round of talks with the government ,	
	3	does not include the energy system or armed forces ,	
	4	members and installations.	

Document 6		lima , 25 oct 89 (efe).
1	1	police have reported that terrorists tonight bombed the embassies of the prc and the soviet union.
2	1	the bombs caused damage
	2	but no injuries.
3	1	a car bomb exploded in front of the prc embassy ,
	2	which is in the lima residential district of san isidro.
4	1	meanwhile ,
	2	two bombs were thrown at a ussr embassy vehicle that was parked in front of the embassy located in orrantia district ,
	3	near san isidro.
5	1	police said the attacks were carried out almost simultaneously and that the bombs broke windows and destroyed the two vehicles.
6	1	no one has claimed responsibility for the attacks so far.
7	1	police sources ,
	2	however ,
	3	have said the attacks could have been carried out by the maoist shining path group or the guevarist tupac amaru revolutionary movement (mrta) group.
8	1	the sources also said that the shining path has attacked soviet interests in peru in the past.
9	1	in july 1989 the shining path bombed a bus carrying nearly 50 soviet marines into the port of el callao.
10	1	fifteen soviet marines were wounded.
11	1	some 3 years ago two marines died following a shining path bombing of a market used by soviet marines.
12	1	in another incident 3 years ago ,
	2	a shining path militant was killed by soviet embassy guards inside the embassy compound.
13	1	the terrorist was carrying dynamite.
14	1	the attacks today come after shining path attacks during
	2	which least 10 buses were burned throughout lima on 24 oct.

Document 7		medellin , 12 apr 90 (dpa).
1	1	medellin mayor juan gomez martinez
	2	once more insisted today on an immediate dialogue between the government and the cocaine traffickers to end the war in
	3	which innocent people are being killed.
2	1	the mayor reiterated his position
	2	when he commented on the attack in
	3	which 20 persons were killed and approximately 100 were injured ,
	4	which was perpetrated yesterday by terrorists on the drug cartel s payroll near itagui municipality.
3	1	the blast from the explosion hit a truck full of policemen ,
	2	who had earlier carried out an anti mafia operation in a peasant village ,
	3	as well as several public buses ,
	4	houses ,
	5	and businesses.
4	1	a police expert said that the car bomb (containing 300 kg of dynamite) was activated by remote control as the elite force patrol passed by.
5	1	according to gomez martinez ,
	2	these actions will only cease
	3	when (president virgilio barco vargas) government puts aside so much prejudice and agrees to the peace talks proposed by the subversives.
6	1	the dialogue will take place sonner or later ,
	2	and it is better to start now insisted the mayor of medellin ,
	3	a city
	4	where the world s most powerful cartel has its base of operations.
7	1	the chief of the criminal gang ,
	2	pablo escobar gaviria ,
	3	ordered the assassination of medellin policemen in retaliation for the government s action against the drug traffickers.
8	1	twenty eight policemen have been killed in this city over the last two weeks.
9	1	according to security organizations ,
	2	yesterday s attack could be escobar gaviria s response to the arrest of adolfo mesa meneses his right hand man.
10	1	mesa meneses was arrested last week by the administrative department of security (das ,
	2	secret police) ,
	3	and the authorities have accused him of assassinating on orders from the mafia leader leftist presidential candidate bernardo jaramillo ossa ,
	4	as well as a chief of police ,
	5	a mayor ,
	6	and several journalists.
11	1	meanwhile ,
	2	hundreds of soldiers ,
	3	policemen ,
	4	and detectives are carrying out large scale operations in the colombian countryside and border areas on the suspicion that escobar gaviria plans to flee the country.
12	1	last night , a television newscast showed photographs of the drug lord that had been recently taken by security organizations at a colombian beach on the pacific coast.
13	1	a police department spokesman confirmed today that we have reliable information on his whereabouts
	2	although he refrained from giving further details.

to be continued...

continued...			
14	1	unofficial sources said that escobar gaviria is somewhere in western antioquia department of	
	2	which medellin is the capital	
	3	where the search operations are concentrated.	
15	1	colombian authorities have also asked for the cooperation of brazil ,	
	2	ecuador ,	
	3	panama ,	
	4	peru ,	
	5	and venezuela to stop the drug traffickers from escaping to those countries.	

Document 8		bogota , 25 sep 89 (afp).
1	1	the police have reported that three terrorist attacks here in bogota and in the caribbean port of cartagena left two dead and three injured.
2	1	the most serious incident took place at the historical beach resort
	2	where a bomb exploded on the 6 th floor of the hilton hotel killing two people and seriously injuring another.
3	1	almost simultaneously another bomb exploded in front of a bank in another part of cartagena ,
	2	partially destroying the building and injuring a guard and a passerby.
4	1	meanwhile ,
	2	in bogota ,
	3	a bomb exploded at a branch of the city s electric energy enterprise in the suburb of quirigua.
5	1	the company s offices were destroyed ,
	2	and approximately 50 neighboring residences were damaged by the blast.
6	1	according to a guard at the company ,
	2	who escaped unhurt ,
	3	two individuals
	4	who were passing by the building placed the explosives and fled.
7	1	the guard said he barely had time to get away from the bomb before it went off.
8	1	guests at the hotel ,
	2	where practically all the windows were destroyed ,
	3	had to be evacuated to other lodgings.
9	1	cartagena had up to now escaped the wave of dynamite attacks that have been taking place in the country s large cities for the past several weeks.
10	1	authorities attribute the attacks to drug traffickers.

D.3 Visualising agreement between coders

The following four pages contain a graphical representation of the agreement between human document coders (the author included) for each document. In order to show this, we have adopted a "bird's eye" view of all sixteen grids (one for each coder) per document. In this view, each horizontal line (to be read from left to right, just as grids are read from top to bottom) represents one coder's grid, with fragments of identically shaded line denoting event coreference. Increasingly lighter shades are used to indicate new events. This means that, for example, if all sixteen coders agreed that the document in question was minimally eventful, the representation used here would be uniformly dark. On the other hand, if all coders agreed that the document was maximally eventful, we would see sixteen parallel lines moving from dark, on the left, to light, on the right, in unison. Furthermore, although representations are uniform in the vertical dimension (as each one contains sixteen lines), they vary in the horizontal dimension depending on the length of the original document that they represent.

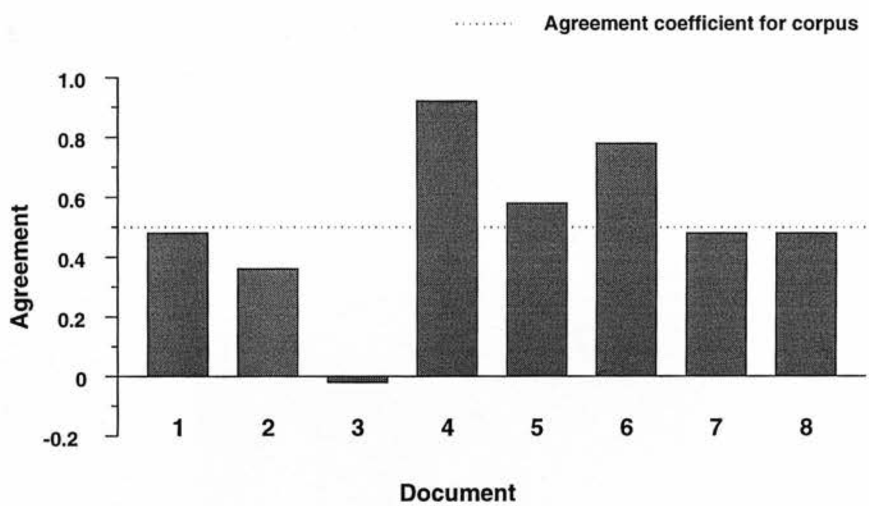


Figure D.1: Individual values of κ

Figure D.1 shows the value of the agreement coefficient κ for each of the eight documents in the correlation corpus. Individual κ values are also provided in numeric form alongside each figure on the following pages.

The κ measure of agreement was introduced in section 9.6. For commentary on the agreement between coders for each document, the reader is referred to the discussion in section 9.11.

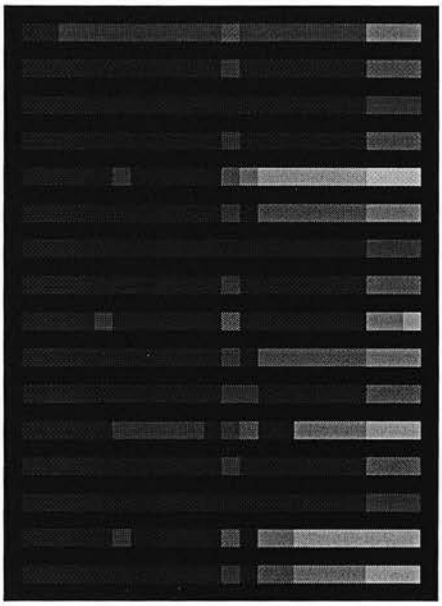


Figure D.2: Document 1 ($\kappa = 0.480126$)

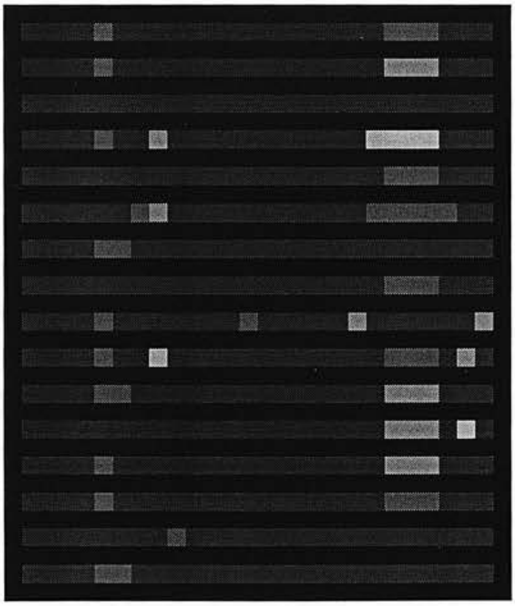


Figure D.3: Document 2 ($\kappa = 0.365195$)

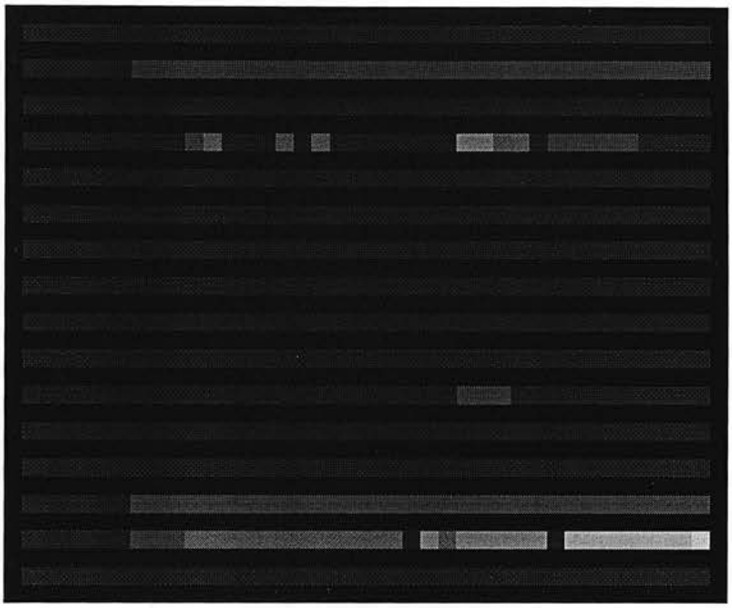


Figure D.4: Document 3 ($\kappa = -0.011159$)

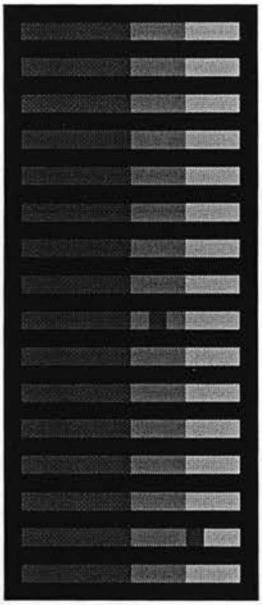


Figure D.5: Document 4 ($\kappa = 0.931016$)

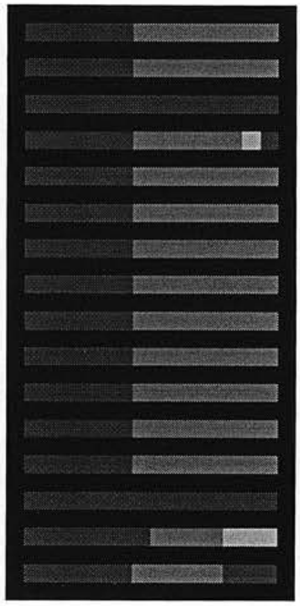


Figure D.6: Document 5 ($\kappa = 0.578093$)

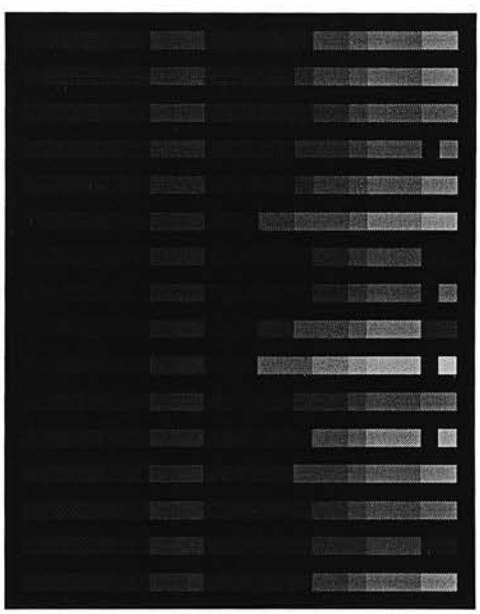


Figure D.7: Document 6 ($\kappa = 0.746846$)

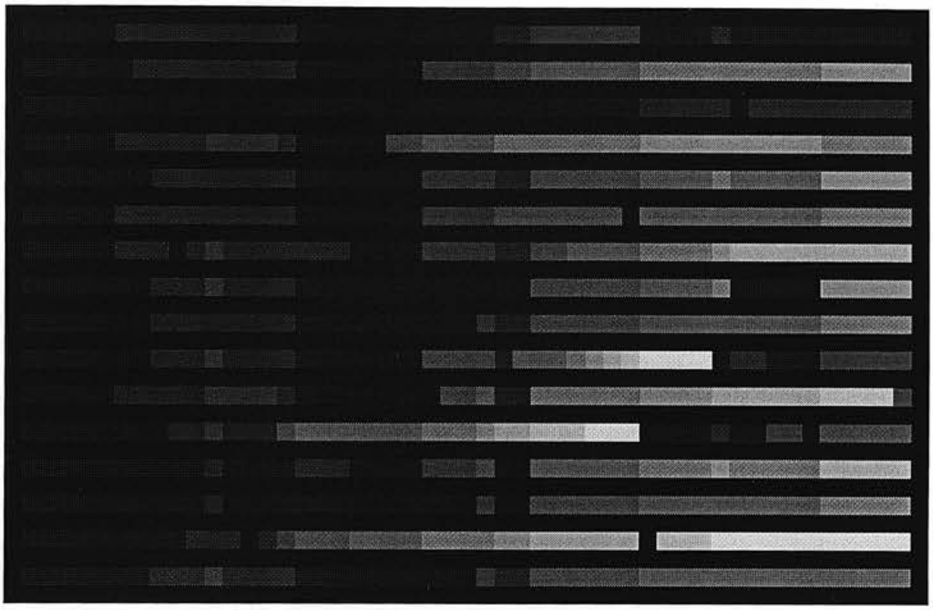


Figure D.8: Document 7 ($\kappa = 0.377622$)

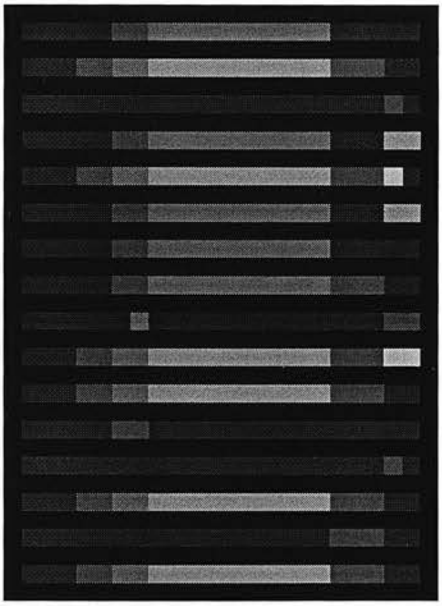


Figure D.9: Document 8 ($\kappa = 0.375408$)

Appendix E

Sample MUC-3/4 documents

E.1 Document with multiple key templates

TST1-MUC3-0001

GUATEMALA CITY, 4 FEB 90 (ACAN-EFE) -- [TEXT] THE GUATEMALA ARMY DENIED TODAY THAT GUERRILLAS ATTACKED THE "SANTO TOMAS" PRESIDENTIAL FARM, LOCATED ON THE PACIFIC SIDE, WHERE PRESIDENT CEREZO HAS BEEN STAYING SINCE 2 FEBRUARY.

A REPORT PUBLISHED BY THE "CERIGUA" NEWS AGENCY -- MOUTHPIECE OF THE GUATEMALAN NATIONAL REVOLUTIONARY UNITY (URNG) -- WHOSE MAIN OFFICES ARE IN MEXICO, SAYS THAT A GUERRILLA COLUMN ATTACKED THE FARM 2 DAYS AGO.

HOWEVER, ARMED FORCES SPOKESMAN COLONEL LUIS ARTURO ISAACS SAID THAT THE ATTACK, WHICH RESULTED IN THE DEATH OF A CIVILIAN WHO WAS PASSING BY AT THE TIME OF THE SKIRMISH, WAS NOT AGAINST THE FARM, AND THAT PRESIDENT CEREZO IS SAFE AND SOUND.

HE ADDED THAT ON 3 FEBRUARY PRESIDENT CEREZO MET WITH THE DIPLOMATIC CORPS ACCREDITED IN GUATEMALA.

THE GOVERNMENT ALSO ISSUED A COMMUNIQUE DESCRIBING THE REBEL REPORT AS "FALSE AND INCORRECT," AND STRESSING THAT THE PRESIDENT WAS NEVER IN DANGER.

COL ISAACS SAID THAT THE GUERRILLAS ATTACKED THE "LA EMINENCIA" FARM LOCATED NEAR THE "SANTO TOMAS" FARM, WHERE THEY BURNED THE FACILITIES AND STOLE FOOD.

A MILITARY PATROL CLASHED WITH A REBEL COLUMN AND INFLICTED THREE CASUALTIES, WHICH WERE TAKEN AWAY BY THE GUERRILLAS WHO FLED TO THE MOUNTAINS, ISAACS NOTED.

HE ALSO REPORTED THAT GUERRILLAS KILLED A PEASANT IN THE CITY OF FLORES, IN THE NORTHERN EL PETEN DEPARTMENT, AND BURNED A TANK TRUCK.

0. MESSAGE: ID	TST1-MUC3-0001
1. MESSAGE: TEMPLATE	1
2. INCIDENT: DATE	02 FEB 90
3. INCIDENT: LOCATION	GUATEMALA: SANTO TOMAS (FARM)
4. INCIDENT: TYPE	ATTACK
5. INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6. INCIDENT: INSTRUMENT ID	-
7. INCIDENT: INSTRUMENT TYPE	-
8. PERP: INCIDENT CATEGORY	TERRORIST ACT
9. PERP: INDIVIDUAL ID	"GUERRILLA COLUMN" / "GUERRILLAS"
10. PERP: ORGANIZATION ID	"GUATEMALAN NATIONAL REVOLUTIONARY UNITY" / "URNG"
11. PERP: ORGANIZATION CONFIDENCE	REPORTED AS FACT / CLAIMED OR ADMITTED: "GUATEMALAN NATIONAL REVOLUTIONARY UNITY" / "URNG"
12. PHYS TGT: ID	"SANTO TOMAS PRESIDENTIAL FARM" / "PRESIDENTIAL FARM"
13. PHYS TGT: TYPE	GOVERNMENT OFFICE OR RESIDENCE: "SANTO TOMAS PRESIDENTIAL FARM" / "PRESIDENTIAL FARM"
14. PHYS TGT: NUMBER	1: "SANTO TOMAS PRESIDENTIAL FARM" / "PRESIDENTIAL FARM"
15. PHYS TGT: FOREIGN NATION	-
16. PHYS TGT: EFFECT OF INCIDENT	-
17. PHYS TGT: TOTAL NUMBER	-
18. HUM TGT: NAME	"CEREZO"
19. HUM TGT: DESCRIPTION	"PRESIDENT": "CEREZO" "CIVILIAN"
20. HUM TGT: TYPE	GOVERNMENT OFFICIAL: "CEREZO" CIVILIAN: "CIVILIAN"
21. HUM TGT: NUMBER	1: "CEREZO" 1: "CIVILIAN"
22. HUM TGT: FOREIGN NATION	-
23. HUM TGT: EFFECT OF INCIDENT	NO INJURY: "CEREZO" DEATH: "CIVILIAN"
24. HUM TGT: TOTAL NUMBER	-

0. MESSAGE: ID	TST1-MUC3-0001
1. MESSAGE: TEMPLATE	3
2. INCIDENT: DATE	02 FEB 90
3. INCIDENT: LOCATION	GUATEMALA: LA EMINENCIA (FARM)
4. INCIDENT: TYPE	ARSON
5. INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6. INCIDENT: INSTRUMENT ID	*
7. INCIDENT: INSTRUMENT TYPE	*
8. PERP: INCIDENT CATEGORY	TERRORIST ACT
9. PERP: INDIVIDUAL ID	"GUERRILLAS"
10. PERP: ORGANIZATION ID	"GUATEMALAN NATIONAL REVOLUTIONARY UNITY" / "URNG"
11. PERP: ORGANIZATION CONFIDENCE	REPORTED AS FACT: "GUATEMALAN NATIONAL REVOLUTIONARY UNITY" / "URNG"
12. PHYS TGT: ID	"FACILITIES"
13. PHYS TGT: TYPE	CIVILIAN RESIDENCE / OTHER: "FACILITIES"
14. PHYS TGT: NUMBER	PLURAL: "FACILITIES"
15. PHYS TGT: FOREIGN NATION	-
16. PHYS TGT: EFFECT OF INCIDENT	SOME DAMAGE: "FACILITIES"
17. PHYS TGT: TOTAL NUMBER	-
18. HUM TGT: NAME	-
19. HUM TGT: DESCRIPTION	-
20. HUM TGT: TYPE	-
21. HUM TGT: NUMBER	-
22. HUM TGT: FOREIGN NATION	-
23. HUM TGT: EFFECT OF INCIDENT	-
24. HUM TGT: TOTAL NUMBER	-

0.	MESSAGE: ID	TST1-MUC3-0001
1.	MESSAGE: TEMPLATE	4
2.	INCIDENT: DATE	02 FEB 90
3.	INCIDENT: LOCATION	GUATEMALA: LA EMINENCIA (FARM)
4.	INCIDENT: TYPE	ROBBERY
5.	INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6.	INCIDENT: INSTRUMENT ID	-
7.	INCIDENT: INSTRUMENT TYPE	-
8.	PERP: INCIDENT CATEGORY	TERRORIST ACT
9.	PERP: INDIVIDUAL ID	"GUERRILLAS"
10.	PERP: ORGANIZATION ID	"GUATEMALAN NATIONAL REVOLUTIONARY UNITY" / "URNG"
11.	PERP: ORGANIZATION CONFIDENCE	REPORTED AS FACT: "GUATEMALAN NATIONAL REVOLUTIONARY UNITY" / "URNG"
12.	PHYS TGT: ID	"LA EMINENCIA FARM" / "FARM"
13.	PHYS TGT: TYPE	CIVILIAN RESIDENCE / OTHER: "LA EMINENCIA FARM" / "FARM"
14.	PHYS TGT: NUMBER	1: "LA EMINENCIA FARM" / "FARM"
15.	PHYS TGT: FOREIGN NATION	-
16.	PHYS TGT: EFFECT OF INCIDENT	PROPERTY TAKEN FROM TARGET: "LA EMINENCIA FARM" / "FARM"
17.	PHYS TGT: TOTAL NUMBER	-
18.	HUM TGT: NAME	-
19.	HUM TGT: DESCRIPTION	-
20.	HUM TGT: TYPE	-
21.	HUM TGT: NUMBER	-
22.	HUM TGT: FOREIGN NATION	-
23.	HUM TGT: EFFECT OF INCIDENT	-
24.	HUM TGT: TOTAL NUMBER	-

0.	MESSAGE: ID	TST1-MUC3-0001
1.	MESSAGE: TEMPLATE	5
2.	INCIDENT: DATE	- 04 FEB 90
3.	INCIDENT: LOCATION	GUATEMALA: PETEN (DEPARTMENT): FLORES (CITY)
4.	INCIDENT: TYPE	ATTACK
5.	INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6.	INCIDENT: INSTRUMENT ID	-
7.	INCIDENT: INSTRUMENT TYPE	-
8.	PERP: INCIDENT CATEGORY	TERRORIST ACT
9.	PERP: INDIVIDUAL ID	"GUERRILLAS"
10.	PERP: ORGANIZATION ID	-
11.	PERP: ORGANIZATION CONFIDENCE	-
12.	PHYS TGT: ID	-
13.	PHYS TGT: TYPE	-
14.	PHYS TGT: NUMBER	-
15.	PHYS TGT: FOREIGN NATION	-
16.	PHYS TGT: EFFECT OF INCIDENT	-
17.	PHYS TGT: TOTAL NUMBER	-
18.	HUM TGT: NAME	-
19.	HUM TGT: DESCRIPTION	"PEASANT"
20.	HUM TGT: TYPE	CIVILIAN: "PEASANT"
21.	HUM TGT: NUMBER	1: "PEASANT"
22.	HUM TGT: FOREIGN NATION	-
23.	HUM TGT: EFFECT OF INCIDENT	DEATH: "PEASANT"
24.	HUM TGT: TOTAL NUMBER	-

0.	MESSAGE: ID	TST1-MUC3-0001
1.	MESSAGE: TEMPLATE	6
2.	INCIDENT: DATE	- 04 FEB 90
3.	INCIDENT: LOCATION	GUATEMALA: PETEN (DEPARTMENT): FLORES (CITY)
4.	INCIDENT: TYPE	ARSON
5.	INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6.	INCIDENT: INSTRUMENT ID	*
7.	INCIDENT: INSTRUMENT TYPE	*
8.	PERP: INCIDENT CATEGORY	TERRORIST ACT
9.	PERP: INDIVIDUAL ID	"GUERRILLAS"
10.	PERP: ORGANIZATION ID	-
11.	PERP: ORGANIZATION CONFIDENCE	-
12.	PHYS TGT: ID	"TANK TRUCK" / "TRUCK"
13.	PHYS TGT: TYPE	TRANSPORT VEHICLE: "TANK TRUCK" / "TRUCK"
14.	PHYS TGT: NUMBER	1: "TANK TRUCK" / "TRUCK"
15.	PHYS TGT: FOREIGN NATION	-
16.	PHYS TGT: EFFECT OF INCIDENT	SOME DAMAGE: "TANK TRUCK" / "TRUCK"
17.	PHYS TGT: TOTAL NUMBER	-
18.	HUM TGT: NAME	-
19.	HUM TGT: DESCRIPTION	-
20.	HUM TGT: TYPE	-
21.	HUM TGT: NUMBER	-
22.	HUM TGT: FOREIGN NATION	-
23.	HUM TGT: EFFECT OF INCIDENT	-
24.	HUM TGT: TOTAL NUMBER	-

E.2 Highly eventful document showing rich locative phrase content

```
TST2-MUC4-0078

CLANDESTINE, 28 NOV 89 (RADIO VENCEREMOS) -- [TEXT] ATTENTION: A REPORT FROM SAN
SALVADOR STATES THAT FIGHTING CONTINUES IN THE NORTHERN SECTOR OF THE CITY. OUR
FORCES ARE FINDING LITTLE RESISTANCE FROM THE SO-CALLED TROOPS BROUGHT FROM THE
INTERIOR OF THE COUNTRY. THE REPORT STATES THAT DURING TODAY'S CLASHES IN CIUDAD
DELGADO, LOS ALPES, GUADALUPE, AND SAN PATRICIO NEIGHBORHOODS, THE TREASURY POLICE
SUFFERED THREE CASUALTIES. WE INFLICTED TWO CASUALTIES ON THE TREASURY POLICE IN
LOS ALPES NEIGHBORHOOD, AND THE 6TH MILITARY DETACHMENT SUFFERED ONE CASUALTY DURING
A CLASH IN SAN PATRICIO NEIGHBORHOOD. IT HAS ALSO BEEN REPORTED THAT FMLN [FARABUNDO
MARTI NATIONAL LIBERATION FRONT] FORCES HAVE SET UP BARRICADES ON THE NORTHERN TRUNK
ROAD AND POWER LINES HAVE BEEN SABOTAGED. AS A RESULT OF THIS SABOTAGE, AN AREA
OF SAN SALVADOR DOES NOT HAVE POWER. WE HAVE ALSO LEARNED THAT TWO TRANSFORMERS
WERE DESTROYED ON KM 10 OF THE NORTHERN TRUNK ROAD, WHERE BARRICADES WERE ALSO SET
UP. THE REPORT ADDS THAT THE ENEMY SHOWED UP 90 MINUTES AFTER THE BARRICADES HAD
BEEN SET UP AND HEAVY FIGHTING RESULTED. THE ENEMY TROOPS SUFFERED FOUR CASUALTIES;
OUR FORCES HAVE REPORTED THAT THEY HAVE NOT ENCOUNTERED PROBLEMS ON THE NORTHERN
TRUNK ROAD. AT THIS HOUR, 1613 [2213 GMT], THE REBEL PRESENCE IN INCREASING ON THE
NORTHERN TRUNK ROAD.

ATTENTION: DURING YESTERDAY'S ATTACK, WITH PEOPLE'S ARTILLERY ON 4TH INFANTRY BRIGADE
```

TROOPS, WE INFLICTED EIGHT CASUALTIES. THESE EIGHT CASUALTIES HAVE BEEN CONFIRMED.

AT NOON TODAY, CLASHES WITH 1ST MILITARY DETACHMENT TROOPS WAS REPORTED IN LA CRUZ HILL, SAN MIGUEL DE MERCEDES JURISDICTION. WE REPEAT: TODAY AT NOON, FIGHTING WAS REPORTED IN LA CRUZ HILL, SAN MIGUEL DE MERCEDES, CHALATENANGO DEPARTMENT.

THIS HAS BEEN THE LATEST REPORT WE HAVE RECEIVED. THERE IS FIGHTING IN CIUDAD DELGADO, LOS ALPES, GUADALUPE, SAN PATRICIO, AND OUR GUERRILLA COLUMNS ARE SPREADING THROUGHOUT THE NORTHERN PART OF THE CITY. IT IS NOW 1615 [2215 GMT] AND THE HIGH COMMAND HAS NOT REACTED. THIS GIVES US AN IDEA OF HOW LOW MORALE IS AMONG THE SOLDIERS IN SAN SALVADOR. THIS NEW BLOW AGAINST THE ENEMY FORCES IN SAN SALVADOR HAS CAUSED THE MORALE OF THE ENEMY FORCES TO DROP EVEN FURTHER.

ATTENTION: WE HAVE JUST LEARNED THAT THERE HAS BEEN A REBEL AMBUSH IN THE NORTHERN SECTOR OF SAN SALVADOR. THE REPORT ADDS THAT AT 0630 [1230 GMT] TODAY, REBEL FORCES AMBUSHED A VEHICLE TRANSPORTING NATIONAL GUARDSMEN. THE AMBUSH OCCURRED ON KM 9.5 ON THE NORTHERN TRUNK ROAD. THE GUARDSMEN WERE HEADING INTO AN AREA WHERE THE FMLN REBELS HAVE SET UP BARRICADES. A POWERFUL GUAZAPA-TYPE MINE AND RIFLES WERE USED IN THE AMBUSH. AT LEAST 10 CASUALTIES, INCLUDING A NATIONAL GUARD OFFICER, WERE INFLICTED. WE REPEAT: A NATIONAL GUARD OFFICER IS INCLUDED AMONG THE CASUALTIES WE INFLICTED DURING THIS AMBUSH CARRIED OUT ON KM 9.5 ON THE NORTHERN TRUNK ROAD.

LIKEWISE, WE HAVE LEARNED THAT FIGHTING CONTINUES IN MEJICANOS, AYUTUXTEPEQUE, CIUDAD DELGADO, AND SOYAPANGO. HEAVY FIGHTING IS ALSO BEING REPORTED IN APOPA AND MARIONA. REBEL FORCES ARE SPREADING OUT THROUGHOUT THE NORTHERN AREA OF SAN SALVADOR AND THEY ARE FINDING LITTLE RESISTANCE. AT 1616 [2216 GMT], THIS HAS BEEN THE REPORT RECEIVED AT RADIO VENCEREMOS. WE ARE RECEIVING REPORTS FROM OUR LIBERATED TERRORITORIES FROM CHALATE [CHALATENANGO], MORAZAN, SAN MIGUEL. GREETINGS TO OUR FORCES IN SAN SALVADOR; ONWARDS FIGHTING COMRADES; OUR PEOPLE ARE SUPPORTING YOU IN SAN SALVADOR.

Because the MUC-3/4 guidelines consider attacks against military targets of the same nationality as the perpetrators to be *guerrilla* attacks rather than terrorist attacks, only one key template is built for this article.

0.	MESSAGE: ID	TST2-MUC4-0078
1.	MESSAGE: TEMPLATE	1
2.	INCIDENT: DATE	28 NOV 89
3.	INCIDENT: LOCATION	EL SALVADOR: SAN SALVADOR (CITY)
4.	INCIDENT: TYPE	ATTACK
5.	INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6.	INCIDENT: INSTRUMENT ID	-
7.	INCIDENT: INSTRUMENT TYPE	-
8.	PERP: INCIDENT CATEGORY	TERRORIST ACT
9.	PERP: INDIVIDUAL ID	"FMLN [FARABUNDO MARTI NATIONAL LIBERATION FRONT] FORCES"
10.	PERP: ORGANIZATION ID	"FMLN" / "FARABUNDO MARTI NATIONAL LIBERATION FRONT"
11.	PERP: ORGANIZATION CONFIDENCE	REPORTED AS FACT / CLAIMED OR ADMITTED: "FMLN" / "FARABUNDO MARTI NATIONAL LIBERATION FRONT"
12.	PHYS TGT: ID	"POWER LINES" "TRANSFORMERS"
13.	PHYS TGT: TYPE	ENERGY: "POWER LINES" ENERGY: "TRANSFORMERS"
14.	PHYS TGT: NUMBER	PLURAL: "POWER LINES" 2: "TRANSFORMERS"
15.	PHYS TGT: FOREIGN NATION	-
16.	PHYS TGT: EFFECT OF INCIDENT	SOME DAMAGE: "POWER LINES" DESTROYED: "TRANSFORMERS"
17.	PHYS TGT: TOTAL NUMBER	-
18.	HUM TGT: NAME	-
19.	HUM TGT: DESCRIPTION	-
20.	HUM TGT: TYPE	-
21.	HUM TGT: NUMBER	-
22.	HUM TGT: FOREIGN NATION	-
23.	HUM TGT: EFFECT OF INCIDENT	-
24.	HUM TGT: TOTAL NUMBER	-

E.3 Uneventful document showing rich locative phrase content

TST2-MUC4-0009

RIO DE JANEIRO (BRAZIL), NO DATE (O GLOBO) -- [TEXT] [CONTINUED] LAWYER CARLOS OLDINEY REPORTED THE GANG AND FEDERAL AGENTS VALDEMIR LOPEZ PRAZERES AND LUIS JOSE DA CONCEICAO AND DETECTIVES PAULO MASSANHYO AND CARLOS ANGEL SERRANO CASTILESSO WERE ARRESTED.

CAMBARALAMAIA'S ORGANIZATION IS JUST ONE OF THE MANY GANGS IN THE REGION THAT DEALS WITH BOLIVIAN COCAINE. ANOTHER GANG IS LED BY GERSON PALERMO, OF UMUARAMA, PARANA STATE, WHO CONTROLS THE EXCHANGE OF COCAINE TO PARAGUAY FOR STOLEN BRAZILIAN CARS. PALERMO OWNS FOUR TWIN-ENGINED CESSNA PLANES AND EMPLOYS MORE THAN 200 MEN. HE ALSO EXCHANGE CARS FOR COCAINE IN BOLIVIA. ALTHOUGH SENTENCED BY THE BRAZILIAN COURTS, PALERMO LIVES IN PEDRO JUAN CABALLERO, PARAGUAY, WHERE HE MAINTAINS BUSINESS CONTACTS WITH THE BOLIVIAN COCA PRODUCERS. PALERMO OBTAINS ETHER AND ACETONE, WHICH ARE ESSENTIAL

FOR REFINING THE DRUG, FROM ARGENTINA.

THE FEDERAL POLICE BELIEVES THAT THE DRUG MIGHT BE REFINED IN MATO GROSSO DO SUL, NEAR SOME RESIDENCES IN THE DOURADOS AND CAMPO GRANDE SECTORS. PALERMO EXCHANGES THE CHEMICALS FOR TONS OF COCAINE PASTE WHICH, IN THE OPINION OF REGIONAL FEDERAL POLICE DELEGATE FRANCISCO VIANA QUEIROZ, IS ALSO PROCESSED IN PARAGUAY. VIANA QUEIROZ MAINTAINS THAT THE DRUG IS REFINED BY A GROUP OF TRAFFICKERS LED BY ARSENIO BENITEZ GONZALEZ, WHO IS KNOWN AS VILHAR AND WHO LIVES IN ASUNCION. THE FEDERAL POLICE BELIEVE THAT VILHAR -- WHO IS BELIEVED TO BE PALERMO'S PARTNER -- IS THE BRAZILIAN TRAFFICKERS' MAIN COCAINE SUPPLIER.

ACCORDING TO THE POLICE, TWO DRUG ROUTES PASS THROUGH MATO GROSSO DO SUL: THE PONTA PORA ROUTE (WHICH IS USED BY PALERMO), AND THE CORUMBA ROUTE, WHICH PASSES THROUGH THE BOLIVIAN TOWN OF PUERTO SUAREZ. MANY TRAFFICKERS COME TO THIS BOLIVIAN TOWN, MOST OF THEM SMALL SUPPLIERS WHO ONLY WANT TO EXCHANGE COCAINE FOR CARS THAT WERE STOLEN IN BRAZIL. ACCORDING TO THE AUTHORITIES, THE CORUMBA ROUTE IS MORE IMPORTANT BECAUSE IT CHANNELS THE SO-CALLED "MINI-TRAFFIC." THIS INVOLVES HUNDREDS OF PEOPLE WHO, TOGETHER, MAKE UP A SIZABLE QUANTITY OF DRUGS EVERY DAY.

THE CARS ARE THE HARD CURRENCY USED IN THESE TRANSACTIONS. MOST OF THEM ARE TAKEN TO BOLIVIA AFTER HAVING BEEN STOLEN IN SAO PAULO. THEY FOLLOW THE SAME ROUTE AS THE BOLIVIAN COCAINE, THAT IS, VIA PUERTO SUAREZ. THE MOST IMPORTANT PART OF THE TRAFFICKING, HOWEVER, IS CARRIED OUT IN PARAGUAY AND BOLIVIA. GERSON PALERMO IS RESPONSIBLE FOR PART OF IT, WHILE THE REST IS SHARED BY HUNDREDS OF BOLIVIAN TRAFFICKERS. PALERMO OBTAINS ETHER AND ACETONE FROM FORMOSA, ARGENTINA. HE TAKES THESE CHEMICALS ABOARD SMALL PLANES TO THE TOWNS OF PILAR OR ENCARNACION, IN PARAGUAY. FROM THERE, ANOTHER GROUP TAKES THE 200-LITER DRUMS ALONG THE CHACO HIGHWAY TO PUERTO GUARANI, IN BOLIVIA. BOLIVIAN TRAFFICKERS TAKE CARE OF THE REST OF THE JOURNEY TO SANTA CRUZ DE LA SIERRA. IN RETURN, PALERMO RECEIVES COCAINE PASTE WHICH IS SHIPPED TO ASUNCION AND, FROM THERE, TO PEDRO JUAN CABALLERO FOR DISTRIBUTION TO THE PROCESSING UNITS. IT IS BELIEVED THAT THE TRAFFICKERS USE SOME RANCHES NEAR PORTO MURTINHO (MATO GROSSO DE SUL), NEAR THE PARAGUAYAN TOWN OF BAHIA NEGRA, TO PROCESS LARGE QUANTITIES OF COCAINE.

NO COCAINE PASTE IS USUALLY AVAILABLE ALONG THE CORUMBA ROUTE. THE DRUG IS REFINED BECAUSE IT IS RELATIVELY EASY TO FIND ETHER AND ACETONE IN CORUMBA, WHERE THE FEDERAL POLICE HAVE FAILED TO REDUCE THE LARGE DEMAND FOR SUCH CHEMICALS IN RETAIL STORES. DOZENS OF DRUGSTORES SELL THOSE PRODUCTS BY THE LITER, AND THIS IS NOT FORBIDDEN BY LAW.

THE AUTHORITIES ESTIMATE THAT PRODUCTION IN THE PUERTO SUAREZ REGION NEARS 300 KG PER WEEK, WHICH IS TRANSPORTED BY RAILROAD, AIR, AND LAND. IN AN ATTEMPT TO EVADE THE FEDERAL POLICE, TRAFFICKERS HAVE TRIED TO TRANSPORT THE DRUG INSIDE THEIR BODIES, IN WINE DEMIJOHNS, AND EVEN IN CHILDREN'S DIAPERS. THESE TRAFFICKERS GENERALLY TRAVEL BY BUS OR BY TRAIN. OVER THE PAST 6 MONTHS THE POLICE CAUGHT DOZENS OF TRAFFICKERS USING THESE METHODS. NOW, HOWEVER, ONLY WITH THE COMPLICITY OF THE POLICE CAN CARS CARRYING BOLIVIAN COCAINE PRODUCED IN THE PUERTO SUAREZ REGION REACH CAMPO GRANDE.

No key template was generated for this document, as it contains no relevant MUC events.

Appendix F

Rule evaluation algorithm and example

This appendix contains a Perl implementation of the rule evaluation algorithm introduced in section 8.4, followed by an illustration of the algorithm applied to the example text as presented throughout the thesis.

F.1 Rule evaluation algorithm

```
#!/usr/local/bin/perl
# rules.pl
# Analyses rules used in interpreting specific clauses and correlates this
# with man and com binary grids to determine which rules, given the whole
# corpus, contribute profitably and which don't.
# Usage: rules.pl [-D doc] module dataset
# e.g.: rules.pl time tst2
# -D [doc] just do DOC

require 'getopts.pl';           # require switches library
do Getopts('D:');

$do_doc = $opt_D;
$module = $ARGV[0];             # module to use
$dataset = $ARGV[1];            # dataset to use
$ruledir = "/hame/jeremyc/phd/$module/output/$dataset/rules";
$comdir = "/hame/jeremyc/phd/eval/sys/grids/$dataset/com/perl/binary";
$mandir = "/hame/jeremyc/phd/eval/sys/grids/$dataset/man/perl/binary";
$idfile = "/hame/jeremyc/phd/eval/sys/idfile";

# $idfile is a file that contains an ordered list of the 128 Contess
# configuration codes
```

```
open(ID,$idfile);
while (<ID>) {
  chop;
  $id[$.] = $_;                 # put codes into a list
}
close(ID);

if ($module eq 'time') {         # if we're examining the TAM, then pattern
  $pattern = 's10\\d\\d\\d\\d0'; # should have LAM and CPAM deselected
}

if ($module eq 'location') {     # likewise for the LAM,
  $pattern = 's01\\d\\d\\d\\d0'; # TAM and CPAM deselected
}

$invpattern = 's00\\d\\d\\d\\d0'; # all modules deselected

opendir(RULES,$ruledir);        # each document in the dataset has a file
@rules = grep(/txt*/,readdir(RULES)); # containing the rules that were used
closedir(RULES);                # for each clause in the document

foreach $file (sort @rules) {   # for each filename in the set
  if ($do_doc && $do_doc ne $file) {
    next;
  }
  print " CONFIG     A    B    RA    RB   C#    BP BITS +/- SCORE(A)  SCORE(B)\n";
  @file = 0;
  open(RFILE,"$ruledir/$file"); # open that file
  while (<RFILE>) {             # for each rule used in the document
    chop;
    $file[$.-1] = $_;           # get the current rule (if any)
    $freq{$_}++;                # and increment the frequency of the rule
  }
  $rr = $.-1;
  close(RFILE);
  open(MFILE,"$mandir/$file") || die "Can't find manual $file!\n";
  open(CFILE,"$comdir/$file") || die "Can't find computer $file!\n";
  while (<MFILE>) {
    chop;
    (@manbits) = split(':');
    $l = (length($_) / 2);
  }
  $normalize = 1/(($l**2 - $l)/2); # normalisation for score changes
  while (<CFILE>) {
    $rl = $rr;
    $orl = $rl;
    chop;
```

```
    (@combits) = split(':');
    $score = 0;
    $c = 0;
    foreach $x (0 .. ($l - 1)) {
      $c++;
      $cur = $orl - $rl;
      if ($file[$cur] ne '[]' && $file[$cur+$c] ne '[]') { # if 2 rules found
        if ($id[$.] =~ /$pattern/) {              # and config is to be examined
          if ($manbits[$x] == $combits[$x]) {    # and bits are the same
            $score{$file[$cur]}+= 1*$normalize; # then increment the scores
            $score{$file[$cur+$c]}+= 1*$normalize; # of both rules
            $act = " +";
          }
          else {                                  # else if bits are different
            $score{$file[$cur]}-= 1*$normalize; # then decrement the scores
            $score{$file[$cur+$c]}-= 1*$normalize; # of both rules
            $act = " -";
          }
        }
        elsif ($id[$.] =~ /$invpattern/) {       # else if module not selected
          if ($manbits[$x] != $combits[$x]) {    # and bits are different
            $score{$file[$cur]}+= 1*$normalize; # then increment the scores
            $score{$file[$cur+$c]}+= 1*$normalize; # of both rules
            $act = "?+";
          }
          else {
            $act = "  ";
          }
        }
        else {
          $act = "  ";
        }
        printf "%8s %3d %3d %4s %4s %3d %4d %3d [%d%d] %2s %9.6f %9.6f\n",
        $id[$.], $cur, $cur+$c, $file[$cur], $file[$cur+$c], $., $x, $rl,
        $manbits[$x], $combits[$x], $act, $score{$file[$cur]},
        $score{$file[$cur+$c]}; # print trace information (as below)
      }
      if ($c == $rl) {          # keep rule pointer right
        $c = 0;
        $rl--;
      }
    }
  }
  close(CFILE);
  close(MFILE);
}

foreach $x (keys %score) {      # print out scores for rules
```

```
  printf "%4s %10.6f %10.6f %3d [%s %s]\n", $x, $score{$x},
  $score{$x}/$freq{$x}, $freq{$x}, $module, $dataset if ($x ne '[]');
}
```

F.2 Output of rule evaluation algorithm applied to example text

The output below contains examples of the score changes assigned to the rules when applied to the example text. As configurations with the LAM or CPAM selected do not affect rule scores, they have been removed below. Columns represent the following: `CONFIG` is the configuration code; `A` and `B` are the *clause* numbers of the text currently being considered; `RA` and `RB` are the rules for these two clauses respectively; `C#` is the configuration number (1 to 128); `BP` is the position of the pointer in the binary strings for the manual and computer generated grids; `BITS` shows what these two bits are respectively; `+/-` shows whether a score increment or decrement is awarded (`?+` denotes a score increment in a case where the module is *not* selected and there is a discrepancy between bits); and finally, `SCORE(A)` and `SCORE(B)` show the current score of rules `RA` and `RB`. Normally only the summary information containing rule scores is output; excerpts from the tracing are included here as an aid to understanding the algorithm.

```
$ rules.pl time demo

 CONFIG    A  B   RA   RB  C#   BP BITS +/- SCORE(A)  SCORE(B)
s1011110   1  2  r11  r16  18   10 [11]  +  0.000673  0.000673
s1011110   1  5  r11   r4  18   13 [11]  +  0.001347  0.000673
s1011110   1  9  r11 r2ii  18   17 [11]  +  0.002020  0.000673
s1011110   2  5  r16   r4  18   21 [11]  +  0.001347  0.001347
s1011110   2  9  r16 r2ii  18   25 [00]  +  0.002020  0.001347
s1011110   5  9   r4 r2ii  18   43 [11]  +  0.002020  0.002020
s1011100   1  2  r11  r16  20   10 [11]  +  0.002694  0.002694
s1011100   1  5  r11   r4  20   13 [11]  +  0.003367  0.002694
s1011100   1  9  r11 r2ii  20   17 [11]  +  0.004040  0.002694
s1011100   2  5  r16   r4  20   21 [11]  +  0.003367  0.003367
s1011100   2  9  r16 r2ii  20   25 [00]  +  0.004040  0.003367
s1011100   5  9   r4 r2ii  20   43 [11]  +  0.004040  0.004040
[...]
s0011110   1  2  r11  r16  50   10 [10] ?+  0.016835  0.016835
s0011110   1  5  r11   r4  50   13 [10] ?+  0.017508  0.016835
```

```
s0011110   1   9  r11 r2ii  50   17 [10] ?+  0.018182  0.016835
s0011110   2   5  r16   r4  50   21 [10] ?+  0.017508  0.017508
[...]
s0000000   5   9   r4 r2ii 128   43 [10] ?+  0.064646  0.053872

 r16   0.053872   0.053872   1 [time demo]
 r11   0.064646   0.064646   1 [time demo]
r2ii   0.053872   0.053872   1 [time demo]
  r4   0.064646   0.064646   1 [time demo]
```